

البنك الشجري النحوي

الرسائل الجامعية ٤

أحمد روبي محمد

الرسائل الجامعية ٤

البنك الشجري النحوي

بناؤه وتوظيفه في إطار تقنيات الذكاء الاصطناعي

أحمد روبي محمد





البنك الشجري النحوي
بناؤه وتوظيفه في إطار تقنيات الذكاء الاصطناعي

الطبعة الأولى
١٤٣٨ هـ - ٢٠١٧م

المملكة العربية السعودية - الرياض
ص.ب ١٢٥٠٠ الرياض ١١٤٧٣
هاتف:٠٠٩٦٦١١٢٥٨٧٢٦٨ - ٠٠٩٦٦١١٢٥٨١٠٨٢
البريد الإليكتروني: nashr@kaica.org.sa

ح/ مركز الملك عبدالله بن عبدالعزيز الدولي لخدمة اللغة العربية، ١٤٣٨هـ.
فهرسة مكتبة الملك فهد الوطنية أثناء النشر
محمد ، احمد روبي
البنك الشجري النحوي: بناؤه وتوظيفه في إطار تقنيات الذكاء الاصطناعي./ احمد روبي محمد
- الرياض، ١٤٣٨هـ
..ص؛ .. سم
ردمك: ٧-٣-٩٠٨٩٥-٦٠٣-٩٧٨
١- اللغة العربية-معالجة البيانات أ. العنوان
ديوي ٤١٨,٢٨٥ ١٤٣٨/٤٦٦٨
رقم الإيداع: ١٤٣٨/٤٦٦٨
ردمك: ٧-٣-٩٠٨٩٥-٦٠٣-٩٧٨

التصميم والإخراج
دار وجوه للنشر والتوزيع
Wojooh Publishing & Distribution House
WOJOOOH
www.wojoooh.com
المملكة العربية السعودية - الرياض
الهاتف:4562410 الفاكس:4561675
للتواصل والنشر:
info@wojoooh.com



المقدمة

مقدمة

الحمدلله رب العالمين، والصلاة والسلام على أشرف المرسلين، وأفصح من نطق بالضاد سيدنا محمد النبي الأميِّ الأمين، وعلى آله وصحبه، ومن اهتدى بهديه إلى يوم الدين، وبعد...

إن التطور الحادث في مجال الذكاء الاصطناعي في علاقته بالعلوم الإنسانية والمعرفية قد عزز من حركة اللسانيات الحاسوبية تجاه الواقع المجتمعي، إذ أصبحت ركنًا أساسيًّا في صناعة مجتمع المعرفة، فضلا عن أن أول عنايتها نمذجة الاستعمال الإنساني للغة ببناء التطبيقات التي تسهم في مجال فهم اللغة المكتوبة والمنطوقة للحاسوب.

وتقتضي تقنيات التعلم الآلي - لا سيما ذات نمط التعلم الموجه - في معالجتها للغة الطبيعية تمثيل اللغة في إطار هندسي يتفق وطبيعة المعطيات الرياضية التي تنطلق منها خوارزميات التدريب في بناء النماذج الإحصائية للغة.

ولما كان النحو أو التركيب هو المسؤول بصفة أساسية عن توفير المعطيات اللازمة للتحليل اللغوي الأعمق، فضلا عن استناده إلى أسس المنطق الرياضي في التمثيل أو التجريد، فقد جاءت الدراسة وعنوانها (**البنك الشجري النحوي: بناؤه وتوظيفه في إطار تقنيات الذكاء الاصطناعي**) محاولة لبناء مدونة موصَّفة توصيفًا نحويًّا للغة

العربية الفصحى المعاصرة، يمكن توظيفها في بناء تطبيقات النمذجة النحوية للغة العربية باستخدام تقنيات الذكاء الاصطناعي.

ونظرًا لتعدد أنماط العربية الفصحى، فإن البحث يقتصر على نمط العربية الفصحى المعاصرة لا سيما لغة الصحافة الإلكترونية؛ لكونها الأكثر انتشارًا واستعمالًا في الواقع اللغوي المعاصر.

ويستمد البحث مادته اللغوية من موقع إسلام أون لاين الإلكتروني في الفترة الزمنية من ٢٠٠٤م إلى ٢٠١٠م، وهي الفترة التي حظي فيها موقع إسلام أون لاين بترتيب متقدمٍ بين مواقع شبكة الإنترنت - من حيث عدد زائريه - وفقًا لتصنيف أليكسا، وذلك مذ إنشاء إدارة الموقع بالقاهرة حتى قبيل الخلاف الحادث بين الشركة المصرية المنتجة للموقع «ميديا إنترناشونال» وجمعية البلاغ القطري المالكة للموقع الرئيسي عام ٢٠١٠م، والتي أدت إلى إغلاق موقع إسلام أون لاين، ونقل تحريره إلى قطر.

ويقوم البنك الشجري النحوي على المنهج الوصفي في التحليل، حيث يعكس أنواع العلاقات التركيبية في النظام الجملي للغة. ويتنوع مجال البحث ما بين الطرق المنهجية في الدرس النحوي التي ترسخت في علم اللغة الحديث مثل لسانيات المدونات Corpus Linguistics التي تعدُّ منطلقًا للتحليل النحوي، وعلم اللغة الإحصائي الذي يناط به وصف البيانات اللغوية وصفًا إحصائيًّا، وعلم اللغة الحاسوبي Computational Linguistics أو معالجة اللغات الطبيعية Natural Language Processing التي تهدف إلى محاكاة اللغة البشرية - عبر تقنيات الذكاء الاصطناعي - بوضع نماذج حاسوبية للإدراك الإنساني.

وسعيًا وراء بلوغ الهدف المنشود من الدراسة، فقد اقتضى الكتاب أن يأتي في تمهيد وخمسة فصول، وخاتمة، وذلك على النحو التالي:

التمهيد: يعرض مفهوم البنك الشجري النحوي، وأهميته على الصعيدين اللغوي والحاسوبي، ثم إرهاصات البنوك الشجرية النحوية بشكل عام، ثم يلي ذلك المبادئ العامة للعنونة اللغوية.

الفصل الأول: بناء المدونة اللغوية. ويتناول مفهوم المدونة اللغوية، واستخداماتها، وإرهاصاتها، ثم اختيار مادة المدونة اللغوية وفق المعايير التي يتم بها اختيار النصوص،

ويلي ذلك تهيئتها بما يتناسب مع النظم الآلية، وأخيرًا تجزئة نصوصها؛ لتكون مؤهلة للعنونة النحوية فيما بعد.

الفصل الثاني: **العنونة بالأقسام الكلامية**: ويتناول تقسيم الكلام وأهميته، وكذلك الحاجة إليه، ثم مفهوم العنونة بالأقسام الكلامية، وأهميتها في تطبيقات المعالجة الآلية للغة الطبيعية بشكل عام، ثم يعرض الفئات الكلامية العربية المختلفة المستخدمة في العديد من التطبيقات، وأخيرا يعرض نماذج من عنونة المدونة اللغوية بالأقسام الكلامية في ضوء اختيار الفئة الكلامية التي تناسب الهدف المنشود.

الفصل الثالث: العنونة بالعلاقات التركيبية. ويتناول تعريف العلاقات التركيبية ودورها في تحديد أبنية الجمل، ثم مفهوم العلاقات التركيبية وأهميتها على الصعيدين اللغوي والحاسوبي، ثم الطرق الإجرائية التي تنطلق منها العنونة التركيبية، وأخيرا يعرض نماذج من عنونة المدونة اللغوية – عينة الدراسة – بالعنونة بالعلاقات التركيبية على المستوى المكوني والوظيفي.

الفصل الرابع: **ترميز البنك الشجري النحوي**: ويتناول مفهوم الترميز بلغة الترميز القابلة للامتداد، ومدى الاستعانة بملف تعريف الوثيقة DTD، ومن ثم إمكانية التحويل أو ترميز البنك الشجري النحوي.

الفصل الخامس: **توظيف البنك الشجري في بناء النماذج النحوية**: ويتناول منهجية بناء النموذج النحوي للغة العربية المعاصرة، موضحًا الخطوات المنهجية لبناء النموذج النحوي للفئات الوظيفية باستخدام تقنيات التعلم الآلي.

الخاتمة والنتائج: وتشتمل على أهم نتائج التي توصل إليها البحث.

وأخيرا، يسرني أن أسدي الشكر مخلصا إلى أساتذتي الكرام الأستاذ الدكتور فريد حيدر أستاذ علم اللغة بكلية دار العلوم جامعة الفيوم، والأستاذ الدكتور محسن رشوان أستاذ الاتصالات والإلكترونيات بكلية الهندسة جامعة القاهرة، والدكتور خالد أبوغالية مدرس علم اللغة بكلية دار العلوم، والأستاذ الدكتور شريف عبده أستاذ تكنولوجيا المعلومات بكلية الحاسبات والمعلومات بجامعة القاهرة، والأستاذ الدكتور ربيع عبد السلام رئيس قسم علم اللغة بكلية دار العلوم جامعة الفيوم، لقاء ما قدموا لي من النصائح والتوجيهات السديدة التي عملت بها. كما أتقدم بالشكر الجزيل إلى

القائمين على مركز الملك عبد الله بن عبد العزيز الدولي لخدمة اللغة العربية ، على ما قدموا لي من إرشاد، ودعم متصل في سبيل الحرص على خروج العمل بأعلى جودة ممكنة، فلهم مني جميعا جزيل الشكر ووافر الامتنان.

أحمد روبي محمد
القاهرة ٢٠ فبراير ٢٠١٧
ahmedaruby@gmail.com

•••

التمهيد

مقدمة:

يضطلع التمهيد بعرض مفهوم البنك الشجري النحوي، وأهميته على الصعيدين اللغوي والحاسوبي، ثم يعرض إرهاصات البنوك الشجرية النحوية بشكل عام، مبينًا المراحل التدريجية للبناء إلى أن اكتمل على النحو الذي نشهده الآن، ثم يلي ذلك المبادئ العامة للعنونة التي يلزم لأي بناء لغوي توصيفي اتباعها سواء أكان على المستوى الصوتي أو الصرفي أو النحوي،.... وقد جاء هذا التمهيد في ثلاث نقاط رئيسية، وذلك على النحو التالي:

١- مفهوم البنك الشجري النحوي وفوائده.

٢- إرهاصات البنوك الشجرية النحوية.

٣- عنونة المدونات اللغوية.

٣-١ لماذا عنونة المدونات اللغوية؟

٣-٢المعايير الأساسية لعنونة المدونات اللغوية.

١ . مفهوم البنك الشجري النحوي، وفوائده.

ما زال العقل البشري في حيرة من أمره إزاء اللغة من حيث وظيفتها وماهيتها ونشاطها داخل المخ الإنساني،[١] إذ تعددت الرؤى ما بين كونها ظاهرة اجتماعية أو نفسية أو ظاهرة عقلية...وأخيرًا ظاهرة بيولوجية أو عصبية، ومع ذلك فقد عجزت هذه الظواهر منفردة عن احتواء إشكالية اللغة، وتقديم تفسيرات لإدراك كنه اللغة.

وفي محاولة اللسانيين الغربيين للإحاطة بتلك الإشكالية اللغوية، «فقد تم استضافة اللغة – نظرًا لتعدد جوانبها – من قبل معظم العلوم الصورية Formal Science التي شملت الإحصاء والرياضيات والمنطق، والعلوم الإخبارية Informative Science من فئة الإنسانيات التي شملت علم النفس، وعلم الاجتماع، وعلم المعرفة، ومن فئة الطبيعيات وقد شملت الكيمياء والطبيعة والفيزياء والبيولوجيا».[٢] وقد أثمر هذا التضافر بين هذه العلوم المعرفية Cognitive sciences تطوّرات هائلة في تجسيد اللغة لا سيما في محاكاتها عبر ميادين الذكاء الاصطناعي التي تقوم على التمثيل الرياضي.

وإذا كان تطبيق القوانين الرياضية والفيزيائية لإدراك الأشياء الممكنة في المحيط الخارجي، فإن وصف اللغة رياضيًّا يجعلنا قادرين على تجسيدها لدى الحاسوب الذي يتخذ الأداء الإنساني نموذجًا له يسعى إلى مناظرته.[٣]

ويقتضي تمثيل اللغة رياضيًّا – من خلال الوصف والاستقراء – الاستناد إلى مدونة لغوية تعكس الواقع اللغوي، بحيث يمكن استنباط القوانين والأحكام التي ترسم الخصائص العامة للظاهرة اللغوية. ومن ثم كانت الحاجة إلى بناء مدونات لغوية توصيفية لا سيما التوصيف النحوي باعتباره من العوامل الأساسية لتوفير المعطيات اللازمة للفهم الأتوماتي للنصوص اللغوية.

(١) علي، (نبيل)، حجازي، (نادية):الفجوة الرقمية «رؤية عربية لمجتمع المعرفة»، عالم المعرفة، ٢٠٠٥، ص٣١٩.

(٢) المصدر السابق، ص٣١٦.

(٣) ينظر: طعمة (عبد الرحمن): ميكانيزمات الإدراك في العقل البشري: دراسة في أساسيات اللغة والوعي من منظور تكنو-عصبي، المؤتمر الدولي الرابع للغة العربية – دبي، ٢٠١٥، ص١٠.

ويمكن أن يعرف البنك الشجري Treebank[١] بأنه مدونة لغوية توصيفية، تتضمن بعض التحليلات النحوية التي تتجاوز بيان مستوى الأجزاء الكلامية إلى الكشف عن العلاقات التركيبية بين هذه الأجزاء.[٢]

١-١ فوائد البنك الشجري النحوي

١- تعد البنوك الشجرية النحوية أساسًا هامًّا لبناء التطبيقات الإحصائية لمعالجة اللغات الطبيعية Natural Language Processing مثل المحللات النحوية Syntactic Parsers، وتطبيقات الترجمة الآلية، وفك اللبس الدلالي للكلمات Word Sense Disambiguation، وغيرها من التطبيقات التي تناظر الأداء الإنساني.[٣]

٢- وتعد – أيضًا – أداة للبحث في الظواهر اللغوية التي تصف الواقع اللغوي، فقد يستخدمها اللغويون النفسيون Psycholinguists في إحصاء ترددات بعض الظواهر التي لها أثر نفسي في فهم اللغة وإنتاجها مثل حصر تكرار أنواع الجمل التكميلية في الواقع اللغوي المنشود، ثم معرفة أي نوع يفضله الإنسان؟. ويستخدمها اللغويون في البحث عن الأنماط التركيبية، والأطر التركيبية Subcategorization Frames للأفعال والأسماء، وكذلك في الكشف عن الخصائص الأسلوبية للنص النثري أو الشعري.[٤]

٣- تمثل البنوك الشجرية النحوية منطلقًا للتَّحقُّق من فرضيات النظريات النحوية القائمة، فضلًا عن إمكانها رسم معالم واضحة لنظام الجملة في اللغة المدروسة، وتحديد خصائص علاقاتها التركيبية، والتي يمكن من خلالها الكشف عن دقائق اللغة وأسرار تجلياتها في الاستعمال اللغوي عبر العصور والأمكنة في مختلف التراكمات اللغوية.

٤- تساعد البنوك الشجرية النحوية على تقييم دِقَّة الأنظمة المُحَوْسَبة وقياسها

(١) قد صك هذا المصطلح – الذي يماثله مصطلح المدونات الموصفة نحويًّا – Parsed Corpora العالم اللغوي جيفري ليتش Geoffrey Leech في الثمانينات من القرن الماضي؛ انطلاقًا من شيوع التمثيل النحوي – وقتذاك – في معظم اللغات بصورة شجرية. ينظر:

Holmboe, H. (2004).Nordisk Sporogteknologi, Nordic Language Technology. Museum Tusculanums forlag.P.99.

(2) Holmboe, H. (2004). Ibid.P.99.

(3) Lüdeling, A and Kyt, A. (2008). Corpus linguistics An International Volum 1. walter de Gruyter GmbH& Co. KG. Berlin, Germany.P.225.

(4) Abeillé, A. (2003). Building and Using Parsed Corpora, Springer. P.xix.

باعتبارها معيارًا ذهبيًّا Gold standard.[(١)]

٢. إرهاصات البنوك الشجرية النحوية

ظهرت فكرة البنوك الشجرية النحوية أو المدوّنات الموصّفة نحويًّا في النصف الثاني من القرن العشرين؛ نتيجة لظهور منهج العنونة اللغوية الذي كان مطلبًا أساسيًّا في تحسس نواميس الظاهرة اللغوية لا سيما الاستعمالات اللغوية في الواقع المعاصر، حيث أعلنت جامعة لَنْد Lund University عن شروعها في بناء مدونة توصيفية للغة السويدية بقيادة اللغوي أولف تلمان Ulf Teleman وزملائه بالقسم، وذلك بهدف استقراء خصائص التركيب اللغوي دون تعسف على الاستعمال.[(٢)]

ثم أعقبت هذه المحاولة عدة محاولات أخرى، منها محاولة اللغوي السويدي إليجراد Ellegrad الذي عمد إلى توصيف مدونة لانكاستر – IBM للغة الإنجليزية توصيفًا نحويًّا، بهدف إكمال تحقيق الشمولية لهذه المدونة من حيث توصيفها لكل المستويات اللغوية،[(٣)] وكذلك محاولة جامعة لانكاستر Lancaster University في عنونة مدونة لانكاستر – أوسلو Lancaster Oslo Bergen (LOB) للغة الإنجليزية البريطانية، إلا أنها تدرجت إلى الآلية في توصيفها.[(٤)]

وبظهور نظرية المعلومات، ونشوء المجالات البينية في إطار اللغويات العامة، اتضحت معالم المنهج، واكتملت أركانه، فكان منطلقًا أساسيًّا في بناء النماذج الإحصائية Statistical Models للغة التي تتم في إطار معالجة اللغة الطبيعية لا سيما أساليب

(١) يعود مصطلح «المعيار الذهبي Gold Standard» إلى المجال الاقتصادي، إذ يشير إلى النظام المالي الذي يتم فيه استعمال الذهب كقاعدة أو معيار لتحديد قيمة العملة وتقييمها، ثم تعددت دلالة المصطلح إلى الإجراءات العلمية التي يمكنها أن تقبل المعايير، فكانت المدونات المعنونة (يدويًّا) معيارًا ذهبيًّا لقياس مدى دقة تطبيقات معالجة اللغة الطبيعية NLP. ينظر:

Wissler , L. & Almashraee, M. & Monett, D. & Paschke, A. (2014). The Gold Standard in Corpus Annotation. In proceedings of the In Proceedings of the 5th IEEE Germany Students Conference 2014. Passau, Germany. P.2.

(2) Lüdeling, A. & Kytö, A. (2008). Ibid.225.

(3) Garsid, R. & Leech, G. & McEnery, T. (2013). Corpus Annotation. Second Published by Routledge. New York USA. P.10.

(4) Garsid, R. & Leech, G. & McEnery, T. (2013). Ibid. P.8.

التعلّم الإحصائي؛ لذلك انطلقت العديد من المشروعات التوصيفية لمختلف اللغات في صورة مشروعات قومية، تدعمها المؤسسات العلمية.

ومن أبرز هذه المشروعات ما تدعمه مؤسسة شبكة البيانات اللغوية LDC[١] بجامعة بنسلفانيا: بنك بنسلفانيا الشجري Penn Treebank للغة الإنجليزية الذي كان بمثابة الشرارة الأولى في توخيه للمبادئ النظرية والتطبيقية، فقد انطلقت منها العديد من البنوك الشجرية الأخرى لكثير من اللغات، وقد أُعلن عنه عام ١٩٩٦م،[٢] وبنك بنسلفانيا الشجري للغة الكورية Korean Treebank الذي أصدر عام ٢٠٠٢م،[٣] وبنك بنسلفانيا الشجري للغة الصينية Penn Chinese Treebank الذي أصدر عن المؤسسة ذاتها عام ٢٠٠٣م،[٤] وبنك بنسلفانيا الشجري للغة العربية Arabic Treebank الذي أصدر عام ٢٠٠٤م،[٥] وبنك براغ الاعتمادي للغة العربية Prague Arabic Dependency Treebank الذي أتاحه معهد اللغويات التطبيقية والصورية بجامعة كارلوفا ببراغ عام ٢٠٠٤م،[٦] وبنك براغ الاعتمادي للغة التشيكية Prague dependency Treebank الذي أعلن عنه عام ٢٠٠٦م. [٧] وبنك كولومبيا الشجري للغة العربية The Columbia Arabic Treebank (CATiB) بدعم من وزارة الدفاع الأمريكية الذي أعلن عنه عام ٢٠٠٩م، والبنك الاعتمادي للنص القرآني Quranic

(1) Linguistic Data Consortium. (1992). https://www.ldc.upenn.edu/

(2) Taylor, A.& Marcus, M. et al., (1996). The Penn Treebank: An Overview. In A. Abeille (Ed.), Treebanks: Building and using parsed corpora (chap. 1). P.5.

(3) Han, ch. & Eon-Suk Ko, Na-Rae. (2002).Penn Korean Treebank: Development and Evaluation. In Preceeding ACL Anthology. WD.

(4) xue, N. & Xia, F. et al., (2003). The Penn Chinese Treebank: Phrase Structure Annotation of a Large Corpus. Journal Natural Language Engineering.

(5) Maamouri, M. & Bies, A.& Buckwalter, T. et al., (2004). The Penn Arabic Treebank: Building a Large-Scale Annotated Arabic Corpus. In NEMLAR Conference on Arabic Language Resources and Tools.

(6) Hajic, J. & Smrz, O. et al., (2004). Prague Arabic Dependency Treebank (PADT): https://catalog.ldc.upenn.edu/LDC2004T23

(7) Cmerjrek, M. & Curin, J. & Havelka, J. (2006). Prague Czech-English Dependency Treebank. In Preceeding ACL Anthology.

Arabic Dependency Treebank (QADT) برعاية جامعة ليدز الذي أصدر عام ٢٠٠٩م.[١]

٣. عنونة المدونات اللغوية

تعنى عنونة المدونة اللغوية[٢] Corpus Annotation بإضافة توصيفات – لغوية أو غير لغوية – إلى مادَّة المدونة اللغوية سواء أكانت في صورةٍ مكتوبةٍ أو منطوقةٍ أو منطوقةٍ في صورةٍ مرئية؛ بغية أن تكون أداةً للبحث والتطويرِ في ميادين معالجة اللغة الطبيعية[٣]، ويتحقَّق ذلك – على مستوى المدوَّنات اللغوية المكتوبة- بإلحاق كلِّ وحدة لغوية برمزٍ أو عدة رموز تشير إليها في النص، بحيث يصبح الوصف ملازمًا للوحدات اللغوية[٤] ومع ذلك يمكن إضافة هذا الوصف بشكلٍ مستقلٍّ عن المادَّة الأصلِيَّة باستخدام أدوات العنْوَنة القادرة على الدمج بين التَّرميز والنَّصِّ الأصليِّ، وكذلك الفصل بينهما.[٥]

نلحظ أن هناك تواردًا لافتًا بين مصطلح «العنونة اللغوية» لدى اللُّغوِيِّين المْحدثِين، وما كان يماهيه الواصف من التَّحليل والتَّأويل للنُّصوص اللُّغوِيَّة عن طريق اليدِّ والعين والقلم عبر العقود الفائتة.[٦] فالوصف هو المنطلق الذي يستشفُّه الواصف في اللُّغة – من خلال تجلِّيَّاتها الكامنة في العقل الإنساني- باللُّغة – أي لغة وصف اللغة Meta Language –، «وذلك بردِّ اللامحدود اللغوي من تجلياته إلى عدد محدود من المقولات اللغوية فعلى سبيل المثال، يتم تصنيف المفردات إلى أقسام الكلم (اسم، فعل، صفة، حرف)، ويتم تنميط الجمل بأنواع تراكيبها وأساليبها (جملة بسيطة وجملة مركبة، جملة

(١) تختلف هذه المشروعات فيما بينها في شيئين أساسيين يخضعان للهدف المنشود من البناء، هما: المعلومات اللغوية التي تقدمها، وطريقة تمثيلها على نحو يتناسب مع طبيعة الغرض البحثي للمشروع.

(٢) هناك مقابل عربي آخر لمصطلح « Annotation" وهو "التحشية"؛ إلا أنني آثرت استخدام المقابل "العنونة" لتداوله بين الباحثين في هذا المجال بحسب اطلاعي.

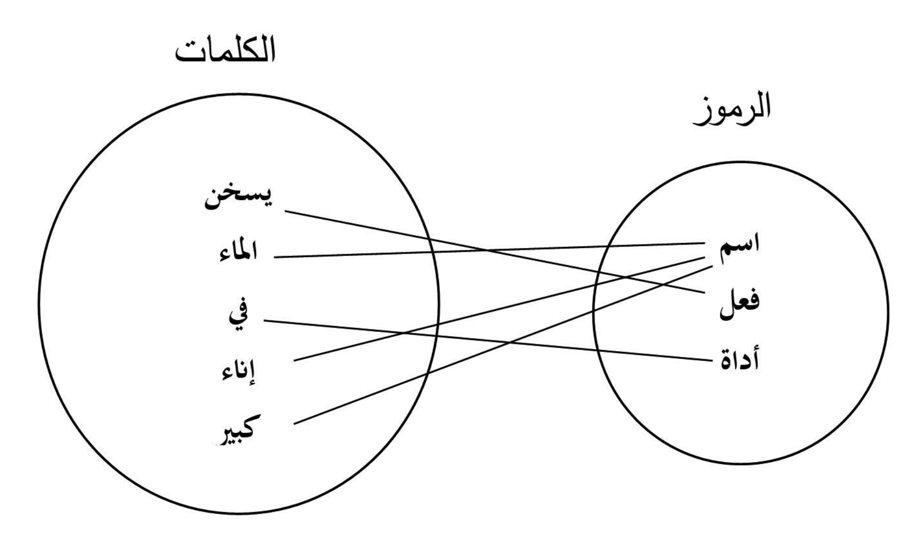
(3) Garsid, R. & Leech, G. & McEnery, T. (2013). Ibid. P.2.

(٤) أما على مستوى المدوَّنات المنطوقة، فتتمّ العنونة بموجب تقسيم الصوت الكلاميِّ إلى وحدات صوتية، ومن ثمَّ تمكينها من إضافة المعلومات حول الظاهرة الصوتية مثل التفخيم أو النبر أو الوقف.

(5) McEnery, T. & Hardie, A. (2012). Corpus Linguistics "Method, Theory and Practice". First Edition. Edinburgh U.P. P.13.

(6) McEnery, T. & Hardie, A. (2012). Ibid. P.13.

خبرية، جملة إنشائية،...)، ويتم تمثيل العلاقات النحوية باستخدام مجموعة من الرموز والدوال النحوية (علاقة الإسناد، علاقة المفعولية، علاقة الوصف، علاقة الإضافة، علاقة الحالية،..)»،[١] وكذلك يتم تعيين المعاني الدلالية بوسمها برموز اصطلاحية تدل بشكلها على معناها. ويوضح الشكل رقم ١ توصيفًا بأقسام الكلام Part Of Speech لمفردات جملة «يسخن الماء في إناء كبير» باستخدام رموز التوصيف Tags[٢] المعبرة عن فئات الكلمات.



الشكل: ١ شكل تمثيلي لتوصيفات الأقسام الكلامية

وتجدر الإشارة إلى أن علماء العربية قد رسموا معطيات النظام الكلي للغة مذ شرعوا في وضع قواعدها في القرن الأول الهجري، إذ تحرُّوا في عملهم أن يردّوا الظواهر اللغوية إلى التشخيص والتجريد، لتوصيف العموم اللغوي في أصول كلية (ناظمة) لمن ينشد أن يتعلّم العربية،[٣] والناظر في الكتب النحوية واللغوية بداية من كتاب

(١) علي (نبيل): اللغة والحاسوب، تعريب، ١٩٨٨م، ص٣٣٦.

(٢) هي رموز توصيفية يتم إدراجها بجوار الكلمات في النص عن طريق المعالجة اليدوية أو الآلية بغرض الوصف والتصنيف.

(٣) ينظر: الموسى (نهاد): العربية نحو توصيف جديد في ضوء اللسانيات الحاسوبية، المؤسسة العربية للدراسات والنشر، الأردن، ٢٠٠٠م، ص ٦١-٦٥.

سيبوية إلى الآن يجد أن محورها الأول تشخيص اللغة وتحليلها بلغة الوصف. ونتلمس ملامح التوصيف في قول ابن جني «الكلام كله ثلاثة أضرب اسم وفعل وحرف جاء لمعنى»[١]، إذ يشخّص المقولات اللغوية، لعلّه يهدي إلى النظر في مكونات الجملة ثم استنباط العلاقات بينها.

ويقتضي التَّوصيف منهجًا يستند إلى مبادئ نظرية تسوّغ التَّحليل والتَّأويل بالاعتماد على الشمول والاختصار في اختيار مجموعة من المعطيات اللغوية Tags set تمثل تلكم المعلومات المنشودة، وعدم التَّناقض في التَّحلِيل بالتعرّف على النظائر في ضوء المعطيات (المقولات) اللغوية.[٢]

٣-١ لماذا عنونة المدونات اللغوية؟

في مضمار السعي لوضع حلولٍ لإشكالية معالجة اللُّغة حاسوبيًّا، أيقن الكثيرون أن حلَّ إشكالية اللُّغة بصورة جادَّةٍ رهن بما يمكن إنجازه على صعيد إشكالية المُخ، والعكس صحيح أيضًا، إذ إن اللُّغة نسقًا معرِفيًّا وفسيولوجِيًّا، تعدُّ من أمضى الوسائل لسبْر أغوار المخ البشريِّ.[٣] وينحي الذكاء الاصطناعيِّ إلى محاكاة هذه المعرفة لدى الإنسان، وسَبْر أغوارها باتباع أساليب منهجيَّةٍ قادرة على التعامل مع هذه المعرفة وطُرُق اكتسابها، ساعيًا إلى تمثيلها في هيئة مخططاتٍ من المفاهيم التي تتداخل مع بعضها البعض لتجسيد البنية اللُّغوِيَّة من الناحية المعرفيَّة، وذلك من قبيل استضافة اللُّغة من قِبَل العلوم الصُّوريَّة التي اشتملت على المنطق والرياضيات والإحصاء ونظريَّة الاحتمالات، وفي إطار عمليَّتي الاستدلال والاستنباط اللذين يقومان عليهما خوارزميات الذكاء الاصطناعيِّ كانت المدوَّنات المُعَنْوَنة أو المُوَصَّفة مجالًا خصبًا ومتطلّبًا حيوِيًّا للتَّعلُّم المُوَجَّه[٤] Supervised learning ؛ لما يستمدّ منها المعطيات التي

(١) ابن جني (أبو الفتح عثمان): اللمع في العربية، تحقيق الدكتور سميح أبو مغلي، دار مجدلاوي للنشر عمان، ١٩٨٨م، ص ١٥.
(2) Kennedy, G. (1998). An Introduction to Corpus Linguistics. Longman. P.220.
(٣) علي (نبيل)، حجازي (نادية): الفجوة الرقمية، ص ٣١٦.
(٤) التعلم الموجه Supervised learning هو أحد الوسائل أو الطرق المستخدمة في التعلم الآلي Machine Learning يستمدّ معطياته من البيانات الموصّفة، لتكون مخرجاته Outputs نابعة من التوصيف المدخل Input، ينظر:
Murphy, Kevin P. (2012). Machine Learning A Probabilistic Perspective. The MIT Press. London, England. P.2..

تنبئ عن الاستدلال والاستنتاج اللُّغويَّيْن.[١]

وثمةَ عديد من تِقنِيَّات التَّعلُّم الآليِّ تستمدّ معطياتها من المُدوَّنات اللغوية المُوصَّفة Tagged Corpora لبناء نماذج إحصائِيَّة لُغوِيَّة من خلال تدريبها على تلك المُدوَّنات التي تعدّ بوصفها حجرَ الأساس في منظومة التَّعلُّم الآليِّ المُوَجَّه لإنتاج التَّطبيقات الحاسُوبيَّة المختلفة للُّغات الطبيعِيَّة، فضلاً عن كونها مُنطَلَقًا لتفسير الظَّواهر اللُّغوِيَّة ونَهْجًا لِلتَّحقُّق من فَرْضِيَّات النَّظرِيَّات اللُّغوِيَّة؛ لذا تبناها اللُّغوِيُّون والحاسُوبِيُّون موضوعًا فتنظيرًا، فتطبيقًا.[٢]

ولما انطلقت المدوَّنات اللُّغوِيَّة التَّوصيفِيَّة في صورة مشروعات علِميَّة وتجاريَّة، ترعاها الجامعات، وتدعمها المؤسَّسات والوزارات الحكومية[٣]، فقد ترسَّخت ملامح أركانها المنهجيَّة، ثم تنامى الاعتناء بها في ضوء المنطلقات لدفع جهود التطوير والبحث في معالجة اللغة الطبيعية؛ لتشمل أنظمة اللغة بمستوياتها اللغوية المختلفة، وذلك في إطار تطبيقها للنظريات اللغوية التي تعنى بالواقع اللغوي ومجتمعه.

٣-٢ المعايير الأساسية لعنونة المدونات اللغوية

ينبغي أن يكون قبولنا للعنونة المفيدة والغنية بالمعلومات مرهونًا إلى حد كبير بتقييمها من حيث توخيها الطرق المنهجية التي تبرز جدوى الوصف، ومدى التزامها بالضوابط النظرية والمنهجية، إذ الضوابط دافعة إلى التقيد بالنتائج المرجوة من الوصف[٤] وبالنظر في إرهاصات الشروح العلمية لعنونات المدوّنات اللغوية، سنجد العديد من الشروحات من الصعب تطبيقها لأنها غير مشروطة بالمنهجية؛ لذا التزم علماء المدوّنات اللغوية بعددٍ من المعايير الأساسية عند شروعهم في عنونة المدوّنات، أوجزها فيما يلي:

١-استخدام طرق العنونة المبسّطة – من ملازمة الوصف للوحدات اللغوية عبر علامة Slash (/) أو (–) Underscore، مثل: كتب/ فعل ماض، كتب–فعل ماض–

(1) Pustejovsky, J. & Stubbs, A. (2012). Ibid. P.21. and Lüdeling, A. & Kytö, M. (2008). Ibid.P. 89.

(2) Indurkhya, N. & Damerau, F. (2010). Handbook of Natural Language Processing. Second Edition.CRC Press.Taylor & Francis Group. London.P.167.

(٣) السعيد (المعتز بالله): مدونة معجم تاريخي للغة العربية، أطروحة دكتوراة، جامعة القاهرة، ٢٠١١م، ص٤.

(4) Garsid, R and Leech, G and McEnery, T. (2013). Ibid. P.6.

بحيث يمكن الفصل بين الترميز والنص مع إمكانية استرداد الصورة الأولية (المادة الخام)؛ للإفادة منها في البحث، والنمذجة اللغوية Language model .[١]

٢-تقتضي الطرق المنهجية لعنونة المدوّنة اللغوية التبصّر في عرض معطيّات النظام اللغوي الذي رسمه علماء اللغة على اختلاف مناهجهم وآرائهم.

٣-ينبغي أن نستشفَّ أولًا المعلومات الأساسية التي سيقدمها التطبيق المنشود، ثم نتخيّر المعلومات (قائمة السمات اللغوية) التي سنستقيها من المدونة اللغوية؛ وذلك للوقوف على منهجية مناسبة للعنونة.

٤-ينبغي أن يحرص صنّاع المدوّنات على التمييز بين الأشباه والنظائر؛ حتى يتسق العمل اتساقًا شديد الإحكام، وفي هذا ضرب من الانسجام بين المفاهيم اللغوية ومدلولاتها، حيث يظهر أثره – فيما بعد – في التطبيقات المنشودة.

٥-قد يستوجب التوصيف الاقتصار على التحليل والتصنيف والوصف، فلا يعتبر – مثلًا – التأويل والحذف والإضمار في الوصف النحوي؛ لأن التوصيف مرتبط بالهدف المنشود.[٢]

٦-تخضع المنهجية – أي منهجية العنونة – لضوابط ومعايير تحددها طبيعة اللغة والهدف المنشود من العنونة، ومن ثمَّ لا تكون هناك منهجية ثابتة باعتبارها معيارًا مطلقًا لعنونة المدوّنات اللغوية، فالمنهجيات تميل إلى التفاوت فيما بينها، فعلى سبيل المثال، تختلف منهجية عنونة أقسام الكلام POS Tagging التي غرضها أن تكون مدخلًا ابتدائيًّا للتحليل النحوي عن التي لها غرض آخر مثل استرجاع المعلومات أو التحليل الدلالي أو غير ذلك.[٣]

٧- ترمّز المدونة الموصّفة لغويًّا بلغة ترميزٍ Markup Language، تقوم على توصيف بيانات النصوص – من خلال مجموعة من العناصر والوسوم يتم نظمها وفق القواعد الكتابية للغة – بحيث يكون لديها قابلية الاستخدام للأنظمة الحاسوبية،

(1) Garsid, R. & Leech, G. & McEnery, T. (2013). Ibid. P.2.

(٢) عاشور(المنصف): بنية الجملة العربية بين التحليل والنظرية، منشورات كلية الآداب بمنوبة – تونس، مجلد ٢، ١٩٩١م، ص٤٦.

(3) Garsid, R. & Leech, G. & McEnery, T. (2013). Ibid. P.7.

وخاصة المتصلة بالشبكة العنكبوتية. وتعتبر لغة الترميز القابلة للامتداد[١] Extensible Markup Language (XML) بمثابة اللغة التوصيفية المثالية، إذ تدعم نظام الحرف الدولي الموحّد للعديد من اللغات الطبيعية.[٢]

وجملة القول أن الالتزام بالضوابط النظرية والمنهجية في عنونة المدونات اللغوية أمر شاق يزيد من صعوبته طبيعة العمل اليدوي في عنونة المدونات التي تتطلب وقتًا وجهدًا كبيرين، إلا أن ثمة أدوات حاسوبية – تمّ تطويرها لهذا الغرض – تساعد محرري المدونات اللغوية على وضع الرموز Tags من خلال مجموعة من المفاتيح – داخل البرنامج– تمثل تلك الرموز، كما تتعرّف –أيضًا– على المفردات المتشابهة التي يحتمل أن يكون لها رمز واحد، مع إتاحة ترميزها مرة واحدة، وغير ذلك من المهام التي تقلل من الوقت والجهد المبذولين.

وتجدر الإشارة إلى إمكانية الإفادة من تقنيات معالجة اللغة الطبيعية – لا سيما المنجزة بالفعل – في عنونة المدونات اللغوية، إذ يمكن استخدامها في تحليل مادة المدونة اللغوية، ثم فحص المخرَج يدويًّا؛ لتقليل نسبة الأخطاء الناتجة عن الآلية. وجدير بالذكر أن استخدام الأدوات والتقنيات المساعدة في العنونة تساعد على تقليل الأخطاء المحتملة الناتجة عن العنونة اليدوية الخالصة.

•••

(١) لغة الترميز القابلة للامتداد XML عبارة عن مجموعة من القواعد لوصف البيانات وتنظيمها، بحيث يستطيع الحاسوب قراءتها، ومن ثم سهولة تنقلها عبر الويب والتطبيقات البرمجية، كما أنها تدعم اليونيكود Unicode الدولي لمختلف اللغات الطبيعية، إذن فهي طريقة لوصف البيانات. ينظر:

Birbeck, M. et al., (2001). Professional XML 2nd Edition. Published by Wrox Press Ltd, the United States. P.5-8.

(٢) السعيد (المعتز بالله): مدونة معجم تاريخي للغة العربية، ص٥١.

الفصل الأول

بناء المدوَّنة اللغويَّة

مقدمة:

يقتضي البناء الشجري النحوي أو العنونة النحوية الاستناد إلى مدونة لغوية، تكون بمثابة النواة الأولى للبناء المنشود؛ لذا تناولت في هذا الفصل - بدايةً - مفهوم المدونة اللغوية، واستخداماتها، وإرهاصاتها، ثم شرعت في اختيار مادة المدونة اللغوية وفق المعايير التي يتم بها اختيار النصوص، بحيث تعكس الواقع اللغوي، مهيئًا نصوصها بما يتناسب مع الأنظمة الآلية، ثم أجريت على هذه النصوص نفسها عملية تجزئة النصوص التي يتم بمقتضاها تقسيم النص إلى جمل ثم إلى وحداتٍ، بحيث تكون مؤهلة للعنونة النحوية فيما بعد. وقد سار هذا الفصل في مبحثين على النحو التالي:

١- المبحث الأول: **مادة المُدوَّنة اللُّغوية Corpus Material**

- المدوَّنة اللغوية، واستخدامها، ومجالات الإفادة منها.
- إرهاصات الإفادة من المدوَّنات اللُّغوية في الدراسات النحوية.
- تحديد مادة المدونة اللغوية عينة الدراسة.
- التحلِيل الإحصائي لمادَّة المدوَّنة اللغوية.
- التهيئة الآليَّة للمدوَّنة اللغوية - عينة الدراسة - قبل إجرائها.

٢- المبحث الثاني: **تجزئة النُّصوص Tokenization**

- مفهوم تجزئة النصوص.
- مستويات تجزئة نصوص المدوَّنة اللُّغوية عينة الدراسة.
- التوحيد الهجائي Orthographic Normalization
- نماذج لتجزئة الوحدات اللغوية في المدونة اللغوية عينة الدراسة.
- التحلِيل الإحصائي لوحدات Tokens المدوَّنة اللُّغوية.

المبحث الأول

مادَّة المُدوَّنة اللُّغويَّة

١ . المدوَّنة اللغوية، واستخدامها، ومجالات الإفادة منها

«تعنى المدونة اللغوية بالبحث في الظواهر اللغوية من خلال مجموعة غير منتظمة من النصوص المكتوبة أو المنطوقة».[١] ومع تطور لسانيات المدونة Corpus linguistics - التي أصبحت قسيما لعلم اللغة الحاسوبي Computational linguistics- أمكن التعامل مع هذه النصوص آليا، والتحكُّم في بياناتها بالإضافة أو الحذف أوالتعديل، فضلًا عن أنها أصبحت حجر الزاوية في بناء النماذج الحاسوبية للغة.[٢]

قد اتَّخذت المدوَّنات اللُّغوية عدة مراحل متعاقبة لاستخدامها، «ففي عام ١٩٥٠م إلى ١٩٦٠م كان الاستخدام الشائع للمدونة اللغوية هو إمكانية إخضاعها للتحليل الإحصائي – باستخدام المفهرسات الآلية Concordances - من جوانب عديدة، منها التعرُّف على شيوع الكلمات وسياقات استعمالها في المدونة المعنية بالدراسة؛ وهو

(1) Kennedy, G. (1998). Ibid P.1. & McEnery, T. & Wilson, A. (2001). Corpus Linguistics "An Introduction". Second Edition. Edinburgh University press.P.29 .

نقلًا عن: السعيد (المعتز بالله): مدونة معجم تاريخي للغة العربية ، ص٥.

(٢) السابق، ص ٥.

مفيد بالطبع لدراسة الظواهر اللغوية على المستويين النحوي والدلالي».[١] وكذلك في احتياجات التعلم.

وما لبث الأمر أن وصل إلى توجيه الأنظمة الحاسوبية الخبيرة إلى فهم اللغات الطبيعية وقدرتها على تمثيل اللغة بمستوياتها كافة من خلال إخضاع الآلة لمتطلبات البحث اللغوي، «وهو ما يشار إليه عادة بتفاعل الإنسان مع الآلة» [٢]، إذ بدأ الباحثون عام ١٩٨٠م بتطويع الخوارزميات الرياضية [٣] للمعالجة اللغوية من خلال تدريبها وتطويرها على الظاهرة اللغوية؛ لتسمحَ للحواسيب بامتلاك خاصية التعلم.

وكان نتاج هذه الفترة تطبيقات التعرُّف الآلي على الكلام Speech Recognition ببناء نماذج لغوية محددة قادرة على التعرُّف على الكلمات/ الصوت – باستخدام المتتابعات اللفظية N-gram [٤] وأدوات ماركوف الخفية HMMs [٥]– من خلال عينات من اللغة المحكيَّة، ثم تطور الأمر في نهاية المنتصف الثاني من القرن العشرين إلى استخدام برمجيات التعلم الآلي Machine Learning [٦] في بناء نماذج إحصائية للترجمة

(1) Pustejovsky, J. & Stubbs, A. (2012). Natural Language Annotation For Machine Learning. Frist Edition.O'Reilly Media. P.10.

(٢) بونيه (الان): الذكاء الاصطناعي واقعه ومستقبله، ترجمة د.علي صبري فرغلي، عالم المعرفة، ١٩٩٣، ص٢٧.

(٣) الخوارزمية الرياضية (Algorithm): هي مجموعة من الخطوات المتسلسلة – الرياضية والمنطقية – لحل مشكلة ما، وتنسب إلى عالم الرياضيات محمد بن موسى الخوارزمي (ت ٢٣٢ هـ). ينظر:

دعبول (موفق)، وآخرون: معجم مصطلحات المعلوماتية، الجمعية العلمية السورية المعلوماتية، دمشق، ٢٠٠١. ص ١٢.

(٤) المتتابعات اللفظية (N-gram): عملية من التتابعات الإحصائية لوحدات النص أو الجملة في ضوء خوارزميات رياضية، وتتوقف N أو عدد المتتابعات على ما تصبو إليه التطبيقات المنشودة، وتسمى أيضا بنمذجة اللغة Language Model. ينظر:

Jurafsky,D. & Martin, J. (2007). Speech and Language Processing: An introduction to natural language processing. Second Edition.P.93.

(٥) نماذج ماركوف الخفية (Hidden Markov Models): هي خوارزمية رياضية تصادفية، تستخدم للتعرف على الخواص الإحصائية لتسلسلات غير ملحوظة، وتعود للعالم أندري ماركوفAndreie Markov في بداية القرن العشرين. ينظر:

Ghahramani, Z. (2001). An Introduction to Hidden Markov Models and Bayesian Networks. International Journal of Pattern Recognition and Artificial Intelligence.P.2.

(٦) التعلم الآلي (Machine Learning): هو امتلاك الآليات الحاسوبية القدرة على التعلم من خلال البيانات، وذلك بتصميم الخوازرميات الرياضية لاستقراء واستنتاج القوانين العامة أو ما تسمى بالتعرف على الأنماط Pattern Recognation من خلال خواص تلكم البيانات. ينظر:

Ethem, A. (2010). Introduction to Machine Learning.The MIT Press Cambridge, Massachustes, London,England. Second Edition. P.2.

الآلية معتمدة على جملة المعطيات التي تمثلها تلك النصوص اللغوية المعنية بالتحليل.[١]

وتعددت مجالات الإفادة من المدونات اللغوية - في دراسة علم اللغة من جانب واللغويات الحاسوبية من جانب آخر- فمنها دراسة اللهجات، وتعليم اللغات، وصناعة المعجمات، والدراسات النحوية، والتحليل الدلالي، والإحصاء اللغوي... وغيرها من متطلبات البحث اللغوي.[٢]

٢. إرهاصات الإفادة من المدونات اللغوية في الدراسات النحوية

لم تكن فكرة استخدام المدونات اللغوية في دراسة جوانب اللغة لاسيما الدراسات النحوية وليدة اللحظة بانصهار الحاسوب في المجتمعات الإنسانية، بل كانت الفكرة[٣] متأصلة في العديد من الدراسات الجادة مذ ما يقرب من بضعة عقود؛ إلا أن حدوث الطفرة المعلوماتية، وتعزيز العلاقة بين اللغة والحاسوب في العصر الحديث شحَذ رؤيتها النظرية، وأرسى مناهجها، وثبت أركانها، وملامحها العلمية.

وقد مرت المدونات اللغوية في هذا المضمار عبر مرحلتين:

٢-١ مرحلة ما قبل ظهور الحاسوب

وتدور محاولات هذه المرحلة حول وصف الظاهرة اللغوية، واستنباط القواعد التي تحكمها من خلال تخيّر عدد ضئيل جدًّا من النماذج النصية أو الصوتية؛ لصعوبة التحليل - آنذاك - الذي يتطلب وقتًا وجهدًا كثيرين، ومع ذلك ينقصه الدقة، ويغلبه الأخطاء المنهجية. ومن أبرز هذه المحاولات:

٢-١-١ محاولة تمثيل قواعد اللغة الإنجليزية

بدأ يسبرسن Jespersen في مطلع النصف الأول من القرن التاسع عشر البحث في جمع مدونة لغوية، مصدرها الصحف والمجلات الإنجليزية؛ لاستخدامها حينذاك

(1) Pustejovsky, J & Stubbes, A. (2012). Ibid.P.13.

(٢) السعيد (المعتز بالله): مدونة معجم تاريخي للغة العربية « معالجة لغوية حاسوبية»، ص ٥.

(٣) يجب الإشارة هنا إلى أن أصل الفكرة والانتفاع بها يعود إلى ما فعلوه النحاة العرب من جمعهم لكلام أهل البادية وأشعارهم، متبعين في ذلك شروطًا وضوابط منهجية مبثوثة في كتبهم، وباستقرائها وضعوا قواعد النحو العربي.

– للمرة الأولى– في الدرس النحوي، فكانت المدونة معطىً مباشرًا لتمثيل قواعد اللغة الإنجليزية كما وردت في كتابه « قواعد النحو الإنجليزي الحديث وفق المبادئ التاريخية» A Modern English Grammar on Historical Principles، ثم تطور الأمر من خلال بعض الدراسات التي تقوم على الوصف النحوي إلى بيان الإحكام النحوي، والتناسق بين العلاقات التركيبية في المدونة اللغوية المحددة زمنيا وآنيًّا. وكان من أهم هذه الدراسات كتاب كروسينجا Kruisinga (١٩٣١-١٩٣٢)، وكتاب بوستما Poustma (١٩٢٦-١٩٢٩).[١]

ثم نشر فريز Fries كتابه «القواعد النحوية للإنجليزية الأمريكية» American English Grammar عام ١٩٤٠م، الذي تضمن أمثلة من الخطابات الرسمية لأعضاء الكونجرس الأمريكي؛ ليعكس واقع الاستخدام اللغوي لقواعد اللغة الإنجليزية. [٢]

وفي عام ١٩٥٩م تطورت هذه المرحلة إلى مرحلة أكثر نضجا، حيث قام راندولف كويرك Randolph Quirk ببناء مدونة جامعة –آنذاك– لاستعمالات اللغة الإنجليزية The Survey Of English Usage (SUE).[٣]

٢-١-٢ محاولة تحليل الجملة العربية

وقد نالت العربية حظًّا من هذه الدراسات، فمنها دراسة المنصف عاشور في كتابه«بنية الجملة العربية بين التحليل والنظرية» عام ١٩٧٧م، حيث استقصى العلاقات التركيبية في نظام الجملة استنادًا إلى مدونة من اللغة العربية، تتضمن ثلاثة نصوص (حدّث أبو هريرة قال لمحمود المسعدي، والمنعرج لمصطفى الفارسي، وخرافات عز الدين المدني)، «وكان ذلك البحث اختبارًا لنظام اللغة العربية وتخصيصًا لقوانينها النحوية ومعانيها المفيدة».[٤]

---

(1) Kennedy, G. (1998). Ibid. P.17.

نقلًا عن: السعيد (المعتز بالله): مدونة معجم عربي معاصر « معالجة لغوية حاسوبية»، أطروحة ماجستير، جامعة القاهرة، ٢٠٠٨م، ص ١٠.

(2) Kennedy, G. (1998). Ibid. P.17.

نقلًا عن: السعيد (المعتز بالله): مدونة معجم عربي معاصر « معالجة لغوية حاسوبية»، ص ١٠.

(٣) السابق، ص ١١.

(٤) عاشور (المنصف): بنية الجملة العربية بين التحليل والنظرية، ص ٢٩١.

وتناول جلال شمس الدين في كتابه «الأنماط الشكلية لكلام العرب» عام ١٩٩٥م فكرة الأنماط اللغوية معتمدًا على نصوص أكثر قربا للاستعمال آنذاك، وتحرّى في عرضه للظاهرة اللغوية على الوصف الآني فقط.

٢-٢ مرحلة ما بعد الحاسوب

لقد أحدثت المدونات الإلكترونية نقلة نوعية في الدراسات اللغوية الحديثة في ظل الإعصار المعلوماتي الجارف الذي كان ملاذًا لأساليب التعلّم الإحصائي، إذ تبلورت مناهجها، وتطورت أدواتها، واكتملت أركانها.

«ففي مطلع السبعينات من القرن المنصرم، وتحديدًا عام ١٩٧٠م، ظهرت فكرة البنوك الشجرية النحوية إبَّان اكتمال أركانها المنهجية، حيث قام أولف تلمان Teleman Ulf وزملاؤه بجامعة لَنْد LUND بأول محاولة جادة لبناء مدونة معنونة بالعلاقات النحوية للغة السويدية، واحتوت هذه المدونة على 300,000 ألف كلمة ما بين مكتوبة ومنطوقة، وهذا يعدُّ إنجازًا عظيمًا -آنذاك- حيث اعتمدوا في صناعتها على الترميز اليدوي لكلا المستويين التمثيلي والوظيفي».[(١)]

«واقتصر الأمر- حينذاك - على اللغة السويدية دون غيرها، حتى اتّضحت معالم المنهج، وتطوَّرت أدواته التحليلية - وذلك في منتصف التسعينات من القرن الماضي - حيث تبنَّت المؤسسات العلمية والتجارية المشروعات اللغوية الكبرى، وتنامى الاعتناء بالبنوك الشجرية النحوية على نطاق واسع في عدة لغات، مع التبصُّر فيما تقتضيه المعالجة الحاسوبية ».[(٢)]

وقد تولت مؤسسة شبكة البيانات اللغوية ببنسلفانيا Linguistic Data Consortium (LDC) – خلال العقدين الماضيين- مشروعات مماثلة للغاتٍ متعددةٍ، كان أبرزها بنك براغ الاعتمادي للغة الإنجليزية والعربية والتشيكوسلوفاكية، وبنك

(1) Lüdeling, A. & Kytö, M. (2008). Corpus Linguistics An International Handbook. Walter de Gruyter.Berlin, Germany.P.226.

(2) Pustejovsky, J. & Stubbs, A. (2012). Ibid. P.226.

بنسلفانيا النحوي للغة العربية والصينية، وقد تعددت مادَّة هذه المشروعات بين النصوص المكتوبة والمنطوقة. [١]

٣. تحديد مادة المدونة اللغوية عينة الدراسة

لما كان الهدف من هذه الدراسة تمثيل العلاقات التركيبية في العربية المعاصرة المكتوبة[٢] لاسيما في الصحافة الإلكترونية المصرية؛ لكونها الأكثر انتشارًا واستعمالًا، وتتوافق إلى حد كبير من حيث البنية التركيبية مع ما تستهدفه تطبيقات معالجة اللغات الطبيعية، فقد روعي عند اختيار نصوص المدونة – عينة الدراسة – عدة معايير:

٣-١ انتماء النصوص للصحافة الإلكترونية :

"العربية المعاصرة لغة مكتوبة تستخدم في مجالات الكتابة، فهي لغة الأدب بمختلف ألوانه ولغة العلم بمختلف فروعه، ولغة الصحافة نمط من هذه الأنماط بل – هي إن أردنا الدقة – تمثل كل الانماط السابقة، إذ تفرد الصحيفة مساحات من صفحاتها للأبواب الثابتة". [٣]

ولغة الصحافة ليست لغة فنية خاصة يمتاز بها مجموعة من الناس، بل هي لغة عامة يتفق من يستطيعون القراءة على فهمها، وما تقدمه لقرائها يجد طريقه ميسّرًا إلى لغتهم حين يكتبون أو حين يتكلمون، فليس غريبًا إذن أن تكون لغة الصحافة أقرب الأنماط السابقة تمثيلًا للخصائص اللغوية التي تميز العربية الآن. [٤]

ومع التطوّر الهائل الذي لحق وسائل الاتصال وتكنولوجيا المعلومات تزايدت أهمية

(1) Linguistic Data Consortium. (1992). https://www.ldc.upenn.edu/

(٢) اللغة المنطوقة ذات حركة دائبة لا يمكن متابعتها، ووضع القواعد التي ترسم خصائصها،... وعليه فاللغة المكتوبة قد نالت حظا أوفر –في الدراسة– من اللغة المنطوقة، «حيث نجد أن المدونات المكتوبة هي الجانب السائد في بناء البنوك الشجرية، ويرجع ذلك الأمر لما تمتاز به من ثبات لاستقراء واستنباط ما تُنْحي عليه النظريات النحوية. ينظر:
Pustejovsky, J & Stubbs, A. (2012). Ibid. P.227.

(٣) عبد العزيز (محمد حسن):لغة الصحافة المعاصرة، دار المعارف، د.ت، ص١١.

(٤) المرجع السابق، ص١٢.

الصحافة الإلكترونية[١] باعتبارها لغة مشتركة بين أغلب فئات المجتمع، فقد أصبحت من أهم وسائل الإعلام المعاصر، وتنشر بين الشرائح المجتمعية كافة بصورة متسارعة، وخاصة بعد أن توفرت شبكة الإنترنت في كل مكان، فضلًا عن أنها تتفوق بعدة سمات كالفورية والتفاعلية، ولا يمكن بأية حال من الأحوال أن تتوافر للصحيفة الورقية المطبوعة. [٢]

٣-١-١ أنواع الصحف الإلكترونية:

٣-١-١-١ الصحف الإلكترونية الكاملة: هي صحف قائمة بذاتها، يتم نشرها وإصدارها إلكترونيًّا فقط، وقد أثبتت هذه الصحف الإلكترونية الكاملة حضورها على المستوى العالمي في مجالات رصد الأحداث، وصناعة الخبر، وتتبع الحدث ونشره في وقت حدوثه، وبذلك سبقت الصحف الورقية بنصف يوم على الأقل.

٣-١-١-٢ النسخ الإلكترونية من الصحف الورقية: وهي نسخ إلكترونية من الصحف التي تصدر بصورة ورقية. وتهدف النسخ الإلكترونية إلى استطلاع الرأي والتفاعل مع القرّاء.

وقد جُمِعَت نصوص المدونة – عينة الدراسة - وفقا للصحف الإلكترونية الكاملة؛ لتعكس الواقع اللغوي في الوقت الراهن من حيث استخدام الأنماط التركيبية المختلفة. وقد استقى البحث مادة المدونة[٣] من موقع إسلام أون لاين Islam On Line [٤]

---

(١) هناك محاولات عديدة لوضع تعريف واضح للصحافة الإلكترونية، ولعل أهمها تعريف الدكتور نواف حازم، حيث عرفها بأنها «هي الصحافة التي تتم عبر طرق إلكترونية، وتعتمد في تكوينها ونشرها على عناصر إلكترونية تستبدل الأدوات التقليدية بتقنيات إلكترونية اتصالية حديثة كالإنترنت، وتستبدل مخرجاتها الورقية بأخرى رقمية.» ينظر: نواف(خالد حازم)، محمد (إبراهيم محمد): الصحافة الإلكترونية، ماهيتها والمسؤولية التقصيرية الناشئة عن نشاطها، مجلة الشريعة والقانون، العدد السادس والأربعين، ٢٠١١م، ص ٢١٣.

(٢) المرجع السابق، ص ٢١٣.

(٣) تمتلك الشركة الهندسية لتطوير النظم الرقمية RDI هذه النصوص بعد شرائها من موقع إسلام أون لاين، وقد أخذتْ هذه النصوص بإذن منها بغرض البحث.

(٤) قد تم إنشاء موقع أون إسلام On Islam باعتباره بديلا عن إسلام أون لاين، ويعمل عليه فريق موقع إسلام أون لاين المصري سابقًا.

في الفترة من ٢٠٠٤ إلى ٢٠١٠.[١]

ووقع هذا الاختيار على إسلام أون لاين لعدة أسباب:

- التزامها بالكتابة العربية الفصحى غير المختلطة إلا نادرًا.
- معظم المحررين العاملين بها من اللغويين؛ وهذا يعكس نقاء المفردات والأساليب اللغوية على الكتابة.
- كثرة زوَّار الموقع يوميًّا مقارنة بالصحف المتناظرة.
- جديَّة الموضوعات والحقائق العلمية التي تتناولها.
- التنوع الموضوعي في كافة المجالات الإنسانية.

٣-٢ المنهجية في اختيار نصوص المدونة اللغوية

لما كانت المدونة اللغوية هي جملة المعطيات التي تخضع للهدف المنشود من صناعتها، فقد تعددت طرق اختيار النصوص بين طريقتين:[٢]

٣-٢-١ الطريقة الأولى:

أسلوب الحصر الشامل Comprehensive Inventory

يستخدم هذا الأسلوب إذا كان الغرض من الدراسة حصر نصوص المجتمع اللغوي التي تماثل الواقع اللغوي برمته. وهذه الطريقة أو الأسلوب يستخدم عادة في صناعات المعاجم التاريخية للغات، أو في الدراسات المسحية ذي الإطار الزمكاني المحدد.

(١) هي الفترة التي حظي فيها موقع إسلام أون لاين بترتيب متقدم بين مواقع شبكة الإنترنت - من حيث عدد زائريه - وفقًا لتصنيف أليكسا، وذلك منذ إنشاء إدارة الموقع بالقاهرة حتى قبيل الخلاف الحادث بين الشركة المصرية المنتجة للموقع «ميديا إنترناشونال» وجمعية البلاغ القطري المالكة للموقع الرئيسي عام ٢٠١٠، والتي أدت إلى إغلاق موقع إسلام أون لاين، ونقل تحريره إلى قطر. ينظر:

wikipedia The free Encyclopedia. (2015). https://goo.gl/m7DD7v

(٢) السعيد (المعتز بالله): مدونة معجم تاريخي للغة العربية « معالجة لغوية حاسوبية»، ص ٣٤-٣٥.

٣-٢-٢ الطريقة الثانية:

نظرية العينات الإحصائية Statistical Sampling Theory

وتعتمد هذه الطريقة على اختيار عينة من النصوص التي تصف الواقع اللغوي سواء أكانت عينة عشوائية أو غير عشوائية [١] بحسب الهدف المنشود. وتستخدم هذه الطريقة في تجسيم خصائص الواقع اللغوي وسماته من خلال عينة من النصوص.

ولما كانت نصوص المدونة اللغوية تمثيلًا لأنماط التركيب الجملي في الاستعمال، ومعطىً مباشرًا لنظام الجملة في اللغة المدروسة، فكانت نظرية العينات الإحصائية أكثر ملاءمة لاستنباط واستقراء الخصائص التركيبية في النحو العربي في لغة (الصحافة الإلكترونية). وقد جُمِعَت نصوص المدونة اللغوية في صورة عينة قصدية لا تخضع لضوابط أو قيود، بحيث تغطي عدة مجالات متنوعة.

٣-٣ تنوع مادة المدونة اللغوية عينة الدراسة

وقد صنّفت مادة المدونة اللغوية- عينة الدراسة - مجاليّا؛ لتوخي أضرب العلاقات التركيبية في بنية الجملة المتمثلة في نصوص المدونة:

- مجال السياسة.
- مجال الاقتصاد.
- مجال الفنون.
- مجال الاجتماع.
- مجال العلوم.
- مجال الرياضة.

---

(١) تنقسم العينات الإحصائية العشوائية إلى:

أ.العينة العشوائية البسيطة Simple Random Sample

ب.العينة العشوائية الطبقية Stratified Random Sample

ج.العينة العشوائية المنتظمة Systematic Random Sample

د.العينة العنقودية أو المتعددة المراحل Cluster Sample

والعينات الإحصائية غير العشوائية إلى:

أ.العينات القصدية Purposive Sample

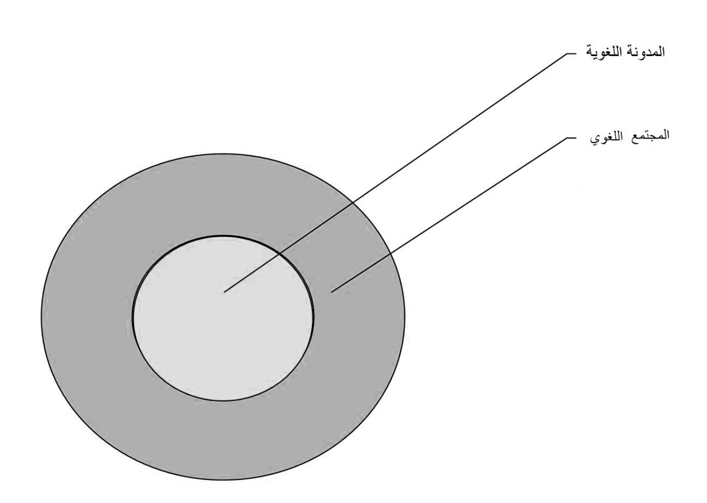
ب. العينات الحصصية Quota Sample

ينظر: خليل (شرف الدين): الإحصاء الوصفي، شبكة الأبحاث والدراسات الاقتصادية، د.ت، ص ١٣.

وتجدر الإشارة إلى أن التنوع في مادة المدونة اللغوية يحقق قدرًا كافيًا من أنماط التركيب الجملي في الاستعمال، والمفردات والأساليب المختلفة، إذ يختلف النمط التركيبي من مجال لآخر. وقد آثرت أن أجعل النصيب الأكبر من النصوص في المجال السياسي؛ ليعكس الواقع اللغوي الراهن الذي تعددت فيه الكتابات والاهتمامات السياسيَّة، ويحقق قدرًا ممكنًا من العلاقات التركيبية المستعملة التي أثَّرت في استعمالها على المجالات الأخرى.

٤. التحليل الإحصائي لمادة المدونة اللغوية

تمثل المدونة اللغوية عينة Sample من مجتمع اللغة غير المحدود Infinite Universe، تحمل خصائصه وصفاته، وتكون مرآةً للواقع اللغوي المعني؛ لكنها تمثل مجتمعًا لغويًّا محدودًا بذاتها Finite Universe، - وهو مجتمع الدراسة الذي يضم مدونات فرعية Sub-Corpora تعدّ في حد ذاتها عينات المجتمع اللغوي المحدود.[١] ويوضح الشكل رقم ٢ تمثيل المدونة اللغوية داخل المجتمع اللغوي.



الشكل:٢ المجتمع اللغوي- المدونة اللغوية (عينة الدراسة).

ولعل التحليل الإحصائي من أهم ما يتصف به العلم – من حيث الانضباط - في التعرُّف على الظاهرة اللغوية وخصائصها، ومدى الانتظام الإحصائي للغة من حيث توزيع المفردات داخل النصوص، والصيغ الصرفية والنحوية، وغيرها.

(١) السعيد (المعتز بالله): مدونة معجم تاريخي للغة العربية «معالجة لغوية حاسوبية»، ص ٥٤.

ويهدف التحليل الإحصائي لمادة المدونة اللغوية إلى وصف بياناتها اللغوية؛ لمعرفة مدى استقرار مصدر البيانات اللغوية من حيث الانتظام الإحصائي لتوزيع المفردات داخل النصوص، وذلك لوجود علاقة إحصائية ثابتة – بين حجم النص اللغوي، وعدد المفردات غير المتكررة فيه – تنم عن انسجام النص اللغوي، وهي العلاقة الإحصائية المعروفة بقانون زيبف Zipf›s law.(١)

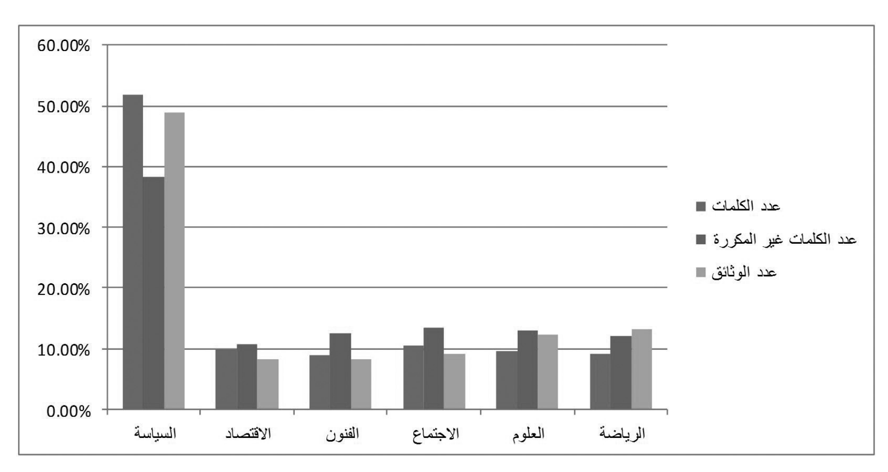
ويعرض الجدول رقم ١ بيانات أو إحصاءات المادة اللغوية عينة الدراسة وفق التصنيف المجالي:

بيانات المدونة اللغة –عينة الدراسة– وفق التصنيف المجالي

الجدول:١ مجالات المدونة اللغوية–عينة الدراسة– وعدد كلماتها ووثائقها ونسبها المئوية

م	المجال	عدد الكلمات	النسبة المئوية ≅	عدد ك.غير المكررة	النسبة المئوية≅	عدد الوثائق	النسبة المئوية≅
1	السياسة	50560	47.8%	13025	34.0%	50	45.45%
2	الاقتصاد	11597	10.96%	4556	11.89%	10	9.09%
3	الفنون	11347	10.73%	5551	14.49%	10	9.09%
4	الاجتماع	11147	10.54%	5117	13.35%	11	10.00%
5	العلوم	11290	10.67%	5401	14.09%	14	12.73%
6	الرياضة	9840	9.30%	4671	12.19%	15	13.64%
-	المجموع	105781	100.00%	38321	100.00%	110	100.00%

(١) علي (نبيل): اللغة العربية والحاسوب، ص٥٣.



الشكل:٣ التمثيل البياني لنسب عدد الكلمات الإجمالية والكلمات غير المكررة وعدد الوثائق وفق التصنيف المجالي

الملاحظ أن أنصبة التوزيع النسبي للكلمات وعدد الوثائق، والكلمات غير المتكررة تميل إلى الثبات أو الاستقرار - في مجالات عينة الدراسة - أي وجود علاقة شبه ثابتة بين حجم النص، وعدد المفردات غير المتكررة فيه، وعدد الوثائق. وهذا ما يؤكد صدق تمثيل العينة المختارة (المدونة اللغوية) من المجتمع اللغوي غير المحدود.[١]

٥. التهيئة الآلية للمدونة اللغوية - عينة الدراسة - قبل إجرائها

يلزم لتهيئة النصوص آليًّا جعلها في صياغة رسميَّة[٢]، بحيث يكون لديها قابلية للتعامل مع الأساليب البرمجية المختلفة للمعالجة الآلية للغة الطبيعية؛ وذلك تمهيدًا لتوظيفها في التطبيقات التي تناظر الأداء الإنساني. وتمرّ تهيئة النص آليًّا بعدة مراحل:

٥-١ تحرير النص Text Editing

يقصد بتحرير النَّصِّ تحويل البيانات النَّصيَّة المتناثرة على الشبكة العنكبوتيَّة أو على صفحات الويب إلى بيانات نصيَّة منتظمة في ملفات نصيَّة Text Document بامتداد

(١) تم حصر مفردات المدونة اللغوية – عينة الدراسة – باستخدام أداة غوَّاص لمعالجة المدونات اللغوية، التي تمَّ تطويرها تحت مظلة مدينة الملك عبد العزيز للعلوم والتقنية.

(٢) صياغة رسمية أي قابليتها للأسس والمبادئ التي تقوم عليها التقنيات الحاسوبية.

dot TXT ، بحيث يسهل التعامل معها بالتعديل أو الحذف. وتعالج هذه الملفات باستخدام المحررات النصية مثل Notepad أو ++Notepad.[١]

٥-٢ حذف المسافات الزائدة Remove Spaces

وتقتضي المعالجة الآلية للغة الطبيعية حذف المسافات الزائدة - في أول السطر أو بين الكلمات أو في نهاية السطر – الناتجة عن عدم الانضباط في عملية إدخال النصوص؛ لأن المسافة الزائدة تعد حرفًا زائدًا، أو كلمة زائدة.

وقد استخدمت الدراسة التعبيرات النمطية Regular Expressions[٢] لحذف المسافات الزائدة في نصوص المدونة اللغوية من خلال النمط الآتي:

```
sed-i 's/^[\ ]*//;s/[\ ]*$//; /^[\ ]*$/d' file.txt
```

٥-٣ توحيد علامات الترقيم Punctuation Normalization

يلزم توحيد الرموز المتماثلة في الشكل مثل علامات الترقيم؛ حتى يتمكن الباحثون في معالجة اللغة العربية آليًّا - لا سيما في بناء النماذج الإحصائية للغة الطبيعية - من حد الضوضاء وتناثر البيانات، حيث تتداخل علامات الترقيم اللاتينية مع علامات الترقيم العربية أثناء إدخال النصوص العربية إلى الحاسوب؛ نتيجة لعدم إدراك مُدخل النصوص إلى الحاسوب؛ ربما للتشابه الشكلي بين هذه العلامات، فعلامات الترقيم في العربية هكذا ؟ ، ؛ : ( ) - .؛ أما الإنجليزية فهكذا ? , :; ( ) -. .

وقد تمت عملية توحيد علامات الترقيم في نصوص مادة المدونة اللغوية من خلال إحصاء كافة الرموز الواردة في المدونة، ثم استبدال علامات الترقيم العربية بعلامات الترقيم اللاتينية الواردة في النصوص.

---

(١) الإصدار الأحدث من محرر النصوص نوت باد Notepad.

(٢) التعبيرات النمطية أو ما تسمى بالريجكس Regex هي تعبيرات نمطية تستخدم لإجراء عمليات إجرائية على النصوص مثل الاستبدال أو الحذف أو البحث،...، بحيث يمكن مطابقة عدد كبير منها بأقل وصف ممكن من التعبير النمطي. وتستخدم في العديد من المنصات البرمجية platforms مثل وجهة سطر الأوامر، أو برامج تحرير النصوص مثل ++Notepad وغيرها.

٥-٤ توحيد الأرقام Numbers Normalization

إن المزج بين الأرقام الهندية والأرقام الإنجليزية- التي أصلها عربية - في النصوص العربية يمثل تحديًا كبيرًا في معالجة اللغة العربية آليا؛ لذا التزمت الدراسة بتوحيد الأرقام بصيغة الأعداد الهندية، "حيث إن الأعداد الهندية أكثر ملاءمة من حيث الشكل وطابعها اليميني مع اتجاه كتابة النصوص العربية".[١]

وقد تمّ توحيد الأرقام في نصوص مادة المدونة اللغوية من خلال عملية استبدال الأرقام الهندية بالأرقام الإنجليزية الواردة في النصوص.

٥-٥ إزالة الكشيدة Tatweel removal

هي زائدة تضاف بين حروف الكلمة، بغرض مساواة النص في الخط العربي، وأحيانًا تضاف في النص دون فائدة، إلا أن وجودها في النص يمثل تحديا كبيرًا في المعالجة الآلية للغة العربية؛ لأنها تؤثر على شكل الكلمة أثناء المعالجة؛ لذلك تمَّ إزالتها من نصوص المدونة اللغوية - عينة الدراسة - من خلال عملية الاستبدال.

وتظهر الكشيدة أو التطويل بين الحروف بهذا الشكل (ـ) كما بالمثال التالي الذي يعرض شكل الكشيدة في الكتابة العربية وتعددها بين الحرف والآخر:

معالجة اللغة الطبيعية

معالجـــــة اللغـــــة الطبيعيـــــة

معالجـــــــــة اللغـــــــــة الطبيعيـــــــــة

معالجـــــــــــة اللغـــــــــــــة الطبيعيـــــــــــة

٥-٦ تشفير النصوص Text Encoding

يقصد بتشفير النصوص تحديد أكواد ثابتة لجميع الحروف، وعلامات التشكيل الأساسية، وعلامات الترقيم، والأرقام، والرموز المستخدمة في النص. وقد تعددت نظم

(١) ينظر: علي (نبيل): اللغة العربية والحاسوب، ص٢١٥.

تشفير النصوص[١]، حيث وضع لكل نظام أكواد خاصة؛ مما أدّى إلى اختلاف نظم التشفير باختلاف بيئات التشغيل وألفبائية اللغة المستخدمة، إلى أن ظهرت نظم تشفير موحدة تدعم العديد من ألفبائيات اللغات الطبيعية، وأهمها نظم التشفير الموحدة Unicod [٢].

وقد اعتمدت الدراسة في تشفيرنصوص المدونة اللغوية على صيغة 8-bit (UTF-8 Unicode Transformation Format) "وهي صيغة تحويل نظام الحروف الدولي الموحّد بقوة ٨ بت"[٣]، الخاضعة لمعايير أيزو ISO التي أقرتها المنظمة الدولية للمعايير القياسية International Organization for Standardization. وتدعم هذه الصيغة العديد من ألفبائيات اللغات الطبيعية، كما تدعم – أيضًا- العديد من نظم التشغيل المختلفة. [٤]

وعلى الرغم من أن هذا الترميز يوفر البنية الأساسية لتهيئة الحروف العربية للمعالجة الآلية، فإنه يعيبه تقارب أشكال بعض الرموز، إذ يقدم عدة تمثيلات لنفس الرمز بصيغة متقاربة، بحيث لا يمكن تمييزها بسهولة، ومثال على ذلك حرف "ك" العربي، وترميزه (U+0643) وحرف "ک" الفارسي، وترميزه (U+06A9)، كلاهما يملكان الشكل الأول لحرف (كـ)."[٥]

ويوضح الشكل رقم ٤ الرموز المستخدمة في صيغة الترميز أو التشفير الموحدة UTF-8 لألفبائيات اللغة العربية والإنجليزية وفقًا لمعايير منظمة ISO .

---

(١) تعتمد تقنيات معالجة اللغات الطبيعية للغة العربية – بصفة رئيسية- على أحد النظم الآتية:
١-نظام الترميز الموحد UTF-8: صيغة التشفير الموحَّدة لجميع الألفبائيات للغات الطبيعية، وجدير بالذكر أن صيغة التشفير UTF-8 اكتسبت صفة العالمية -عمليًا- إذ تمثل حوالي ٨٢٪ من الصفحات المشفرة بهذه الصيغة على الشبكة العنكبوتية.
٢- نظام ترميز Code page 1256 (CP-1256): صيغة تشفير خاص لنظام Windows تدعم اللغات الشبيهة للألفبائية العربية مثل اللغة العربية، الفارسية، والأردية، إلا أنها لا تخضع لمعايير المنظمة الدولية أيزو ISO.
٣- نظام ترميز ISO-8859-6 :صيغة تشفيرتستخدم في ترميز الحروف العربية، وقد طورتها منظمة المقاييس العربية. ينظر:
Habash , N. (2010). Introduction to Arabic Natural Language Processing, A Publication in the Morgan & Claypool Publishers series.P.19.

(٢) أصدرت منظمة Unicode الرمز العالمي الموحد الذي يحوي ٦٥٥٣٦ حرفا، وقد تم تخصيص حوالي ٣٤٠٠٠ حرف منها للغات الحية. ينظر: خضر (محمد ذكي): الحروف العربية والحاسوب، مجلة مجمع اللغة الأردني-عمان، ١٩٩٦م، ص٧.

(٣) أي طول مساحة تخزين الحرف الواحد في ذاكرة الحاسوب يماثل ٨ خانات من العدد الثنائي ٠ و ١ – وهي طول مساحة وحدة البايت التي تحتوي على٨ بت- وتتكون وحدة البايت ٢ ^٨ =٢٥٦ احتمال مختلف يخزن البت إحداها من ٠٠٠٠٠٠٠٠ إلى ١١١١١١١١.

(٤) السعيد (المعتز بالله): مدونة معجم تاريخي للغة العربية « معالجة لغوية حاسوبية»، ص ٥١.

(5) Habash , N. (2010). Ibid. P.20.

	00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F
00	NUL 0000	STX 0001	SOT 0002	ETX 0003	EOT 0004	ENQ 0005	ACK 0006	BEL 0007	BS 0008	HT 0009	LF 000A	VT 000B	FF 000C	CR 000D	SO 000E	SI 000F
10	DLE 0010	DC1 0011	DC2 0012	DC3 0013	DC4 0014	NAK 0015	SYN 0016	ETB 0017	CAN 0018	EM 0019	SUB 001A	ESC 001B	FS 001C	GS 001D	RS 001E	US 001F
20	SP 0020	! 0021	" 0022	# 0023	$ 0024	% 0025	& 0026	' 0027	(0028	) 0029	* 002A	+ 002B	, 002C	- 002D	. 002E	/ 002F
30	0 0030	1 0031	2 0032	3 0033	4 0034	5 0035	6 0036	7 0037	8 0038	9 0039	: 003A	; 003B	< 003C	= 003D	> 003E	? 003F
40	@ 0040	A 0041	B 0042	C 0043	D 0044	E 0045	F 0046	G 0047	H 0048	I 0049	J 004A	K 004B	L 004C	M 004D	N 004E	O 004F
50	P 0050	Q 0051	R 0052	S 0053	T 0054	U 0055	V 0056	W 0057	X 0058	Y 0059	Z 005A	[005B	\ 005C	] 005D	^ 005E	_ 005F
60	` 0060	a 0061	b 0062	c 0063	d 0064	e 0065	f 0066	g 0067	h 0068	i 0069	j 006A	k 006B	l 006C	m 006D	n 006E	o 006F
70	p 0070	q 0071	r 0072	s 0073	t 0074	u 0075	v 0076	w 0077	x 0078	y 0079	z 007A	{ 007B	| 007C	} 007D	~ 007E	DEL 007F
80																
90																
A0	NBSP 00A0				¤ 00A4								، 060C	- 00AD		
B0												؛ 061B				؟ 061F
C0		ء 0621	آ 0622	أ 0623	ؤ 0624	إ 0625	ئ 0626	ا 0627	ب 0628	ة 0629	ت 062A	ث 062B	ج 062C	ح 062D	خ 062E	د 062F
D0	ذ 0630	ر 0631	ز 0632	س 0633	ش 0634	ص 0635	ض 0636	ط 0637	ظ 0638	ع 0639	غ 063A					
E0	ـ 0640	ف 0641	ق 0642	ك 0643	ل 0644	م 0645	ن 0646	ه 0647	و 0648	ى 0649	ي 064A	ً 064B	ٌ 064C	ٍ 064D	َ 064E	ُ 064F
F0	ِ 0650	ّ 0651	ْ 0652													

الشكل:٤ الرموز المستخدمة في صيغة التشفير أو الترميز UTF-8 (١)

يلاحظ أن لكل حرف أو علامة ترقيم أو رمز كودًا يتكون من أربعة أرقام أو أرقام وعلامات، ويتم تكوين هذه الرموز الرباعية لكل حرف وفقًا للثنائية الرأسية والأفقية التي وضعتها المنظمة الدولية للمعايير.(٢)

(1) Unicode UTF-8. (2015). https://msdn.microsoft.com/en-us/goglobal/cc305172
هذه الرموز من المهم جدا أن يكون اللغوي على دراية بها، إذ قد يواجه بعض المشكلات في التشفير التي لاتظهر بشكل مباشر أثناء معالجته لبعض النصوص.

(٢) يمكن معرفة الشفرة الخاصة بكل حرف أثناء كتابتك في برنامج أوفيس ورد Office Word من خلال تظليل الحرف، ثم الضغط على زري ALT+X.

وقد استعانت الدراسة بمحرر النصوص ++Notepad [١] لتشفير نصوص المدونة اللغوية بصيغة UTF-8.

٥-٧ التمثيل الكتابي Orthographic Transliteration

لجأ العديد من الباحثين في معالجة اللغة العربية آليًّا إلى استخدام التمثيل الكتابي[٢]؛ تجنبًا للتحدِّيَّات الناتجة عن ترميز اليونيكود Unicode. «وهو عملية نقل هجائي من لغة ما إلى هجاء لغة أخرى، وفقا لمعيار أنظمة كتابتها.» [٣]

وتستخدم عدة أنظمة للتمثيل الكتابي في معالجة اللغة العربية آليًّا، أشهرها نظام باكولتر الكتابي (Transliteration Buckwalter) الذي يتبع الترميز المعياري للحروف العربية، بحيث يقابلها أي الحروف والعلامات العربية رموزٌ – في أغلبها – إنجليزية؛ لتكون أكثر موثوقية في اكتشاف أخطاء ترميز اليونيكود.'' [٤]

ويعرض الجدول رقم ٢ رموز نظام باكولتر الكتابي بمقابلاتها العربية:

(١) للمطور الفرنسي «دون هو» Don Ho الذي اعتمد على مكونات محرر Scintilla editor المحرر بلغة السي بلس بلس البرمجية ++C في تطوير هذا المحرر؛ ليعمل على بيئة التشغيل ويندوز. يمكن تحميله من خلال هذا الرابط: Notepad++. (2015). http://notepad-plus-plus.org/

(٢) وهناك مقابل آخر لمصطلح « Orthographic Transliteration'' وهو ''النقحرة''.

(3) Habash , N. (2010). Ibid. P.20.

(4) Habash , N. (2010). Ibid. P.21.

Arabic	Unicode Letter Name	HSB	Buckwalter			CP-1256	ISO-8859-6	Unicode
			base	xml	safe			
ء	Hamza	'	'	'	C	C1	C1	0621
آ	Alef Madda Above	Ā	|	|	M	C2	C2	0622
أ	Alef Hamza Above	Â	>	O	O	C3	C3	0623
ؤ	Waw Hamza Above	ŵ	&	W	W	C4	C4	0624
إ	Alef Hamza Below	Ǎ	<	I	I	C5	C5	0625
ئ	Yeh Hamza Above	ŷ	}	}	Q	C6	C6	0626
ا	Alef	A	A	A	A	C7	C7	0627
ب	Beh	b	b	b	b	C8	C8	0628
ة	Teh Marbuta	ħ	p	p	p	C9	C9	0629
ت	Teh	t	t	t	t	CA	CA	062A
ث	Theh	θ	v	v	v	CB	CB	062B
ج	Jeem	j	j	j	j	CC	CC	062C
ح	Hah	H	H	H	H	CD	CD	062D
خ	Khah	x	x	x	x	CE	CE	062E
د	Dal	d	d	d	d	CF	CF	062F
ذ	Thal	ð	*	*	V	D0	D0	0630
ر	Reh	r	r	r	r	D1	D1	0631
ز	Zain	z	z	z	z	D2	D2	0632
س	Seen	s	s	s	s	D3	D3	0633
ش	Sheen	š	$	$	c	D4	D4	0634
ص	Sad	S	S	S	S	D5	D5	0635
ض	Dad	D	D	D	D	D6	D6	0636
ط	Tah	T	T	T	T	D8	D7	0637
ظ	Zah	Ď	Z	Z	Z	D9	D8	0638
ع	Ain	ς	E	E	E	DA	D9	0639
غ	Ghain	γ	g	g	g	DB	DA	063A
ف	Feh	f	f	f	f	DD	E1	0641
ق	Qaf	q	q	q	q	DE	E2	0642
ك	Kaf	k	k	k	k	DF	E3	0643
ل	Lam	l	l	l	l	E1	E4	0644
م	Meem	m	m	m	m	E3	E5	0645
ن	Noon	n	n	n	n	E4	E6	0646
ه	Heh	h	h	h	h	E5	E7	0647
و	Waw	w	w	w	w	E6	E8	0648
ى	Alef Maksura	ý	Y	Y	Y	EC	E9	0649
ي	Yeh	y	y	y	y	ED	EA	064A

جدول:٢ أنظمة التمثيل الكتابي المختلفة للأحرف العربية[١]

(1) Habash , N. (2010). Ibid. P.25.

Arabic	Unicode Letter Name	HSB	Buckwalter			CP-1256	ISO-8859-6	Unicode
			base	xml	safe			
ـً	Fathatan	ã	F	F	F	F0	EB	064B
ـٌ	Dammatan	ũ	N	N	N	F1	EC	064C
ـٍ	Kasratan	ĩ	K	K	K	F2	ED	064D
ـَ	Fatha	a	a	a	a	F3	EE	064E
ـُ	Damma	u	u	u	u	F5	EF	064F
ـِ	Kasra	i	i	i	i	F6	F0	0650
ـّ	Shadda	~	~	~	~	F8	F1	0651
ـْ	Sukun	.	o	o	o	FA	F2	0652
ـٰ	Dagger Alef	á	`	`	e			0670
ٱ	Alef Wasla	Ä	{	{	L			0671
ـ	Tatweel	_	_	_	_	DC	E0	0640
،	Comma	,	,	,	,	A1	AC	060C
-	Soft Hyphen	-	-	-	-	AD	AD	00AD
؛	Semicolon	;	;	;	;	BA	BB	061B
؟	Question Mark	?	?	?	?	BF	BF	061F
پ	Peh	p	P	P	P	81		067E
چ	Tcheh	c	J	J	J	8D		0686
ڤ	Veh	v	V	V	B			06A4
گ	Gaf	g	G	G	G	90		06AF

وما يهمنا – في هذا الصدد- من المعلومات الواردة في هذا الجدول هي الحروف العربية وعلامات التشكيل والرموز المستعملة في النصوص فحسب، ومقابلات ذلك من رموز باكولتر الأساسية المستعملة في نظام التحويل.

وقد استعانت الدراسة بأداة تحويل رموز باكولتر إلى الرموز العربية script buckwalter2unicode [1] للمطور البريطاني أندرو روبرتس Andre Roberts، وذلك لتحويل نصوص المدونة اللغوية – عينة الدراسة- إلى نظام باكولتر الكتابي. وقد حُرِّرت

(١) قمت بتعديل طفيف داخل هذه الأداة لعكس عملية التحويل من رموز باكولتر إلى الرموز العربية؛ لتتوافق مع ما يستهدف البحث.

هذه الأداة بلغة البايثون البرمجية Python[١]، كما تدعم العديد من بيئات التشغيل المتعددة.

٥-٨ إعادة تسمية الملفات Files Rename

وتتجلَّى أهميَّة توحيد صياغة تسمية الملفات؛ لتهيئتها للعتاد البرمجي، بحيث تسهل قراءتها لدى أنظمة التشغيل المختلفة أثناء المعالجة الآلية. وكانت التسمية المقترحة لملفات النصوص بالترتيب الآتي: اسم البناء_اسم مصدر المادة_اسم المجال_التاريخ_رقم الملف. الامتداد.

ويوضح الشكل رقم ٥ تسمية ملفات نصوص المدونة وفق التسمية المقترحة:

SATB_IOL_Spo_2015_0002.txt	10/18/2014 8:06 PM	Text Document	1 KB
SATB_IOL_Spo_2015_0003.txt	10/18/2014 8:06 PM	Text Document	1 KB
SATB_IOL_Spo_2015_0004.txt	11/14/2014 3:43 AM	Text Document	2 KB
SATB_IOL_Spo_2015_0005.txt	11/14/2014 3:25 AM	Text Document	1 KB
SATB_IOL_Spo_2015_0006.txt	11/14/2014 3:43 AM	Text Document	3 KB
SATB_IOL_Spo_2015_0007.txt	11/14/2014 3:43 AM	Text Document	2 KB
SATB_IOL_Spo_2015_0008.txt	10/18/2014 8:07 PM	Text Document	1 KB
SATB_IOL_Spo_2015_0009.txt	11/14/2014 3:43 AM	Text Document	2 KB
SATB_IOL_Spo_2015_0010.txt	11/14/2014 4:00 AM	Text Document	1 KB
SATB_IOL_Spo_2015_0011.txt	11/14/2014 4:00 AM	Text Document	1 KB

شكل:٥ تسمية ملفات النصوص

حيث تعنى SATB اسم البناء المقترح Syntactic Arabic TreeBank، وذلك مقاربة من عنوان البناء (بنك شجري نحوي)، أما IOL فتعني اسم مصدر المادة (إسلام أون لاين Islam On Line)؛ بينما تعني Spo اسم المجال الرياضي Sport.

وقد استعانت الدراسة بأداة بالك المساعدة لإعادة التسمية Bulk Rename[٢] Utility في تسمية ملفات نصوص المدونة اللغوية، إذ يمكن من خلالها تسمية عدد كبير من الملفات في وقت واحد.

(١) لغة البايثون Python لغة برمجية، تم تطويرها من قبل جويدو فان رزوم في أواخر الثمانينات من القرن المنصرم في مركز CWI بأمستردام، وتعد لغة البايثون من لغات المستوى العالي، وقد استخدمت مؤخرًا بشكل واسع في العديد من المجالات، لا سيما في معالجة اللغات الطبيعية.

(٢) للمطور الأمريكي جيم يلشير Jim Willsher؛ وهي من ضمن البرمجيات المفتوحة المصدر Open Source، وتعمل على بيئة التشغيل ويندوز Windows . ويمكن تحميلها من خلال الرابط التالي:
Bulk Rename Utility. (2015). http://www.bulkrenameutility.co.uk/Download.php

المبحث الثاني

تجزئة النُّصوص

١ . مفهوم تجزئة النصوص Tokenization

يعتقد الباحثون في العلوم المعرفية أن إدراك المخ البشري للنص اللغوي وفهمه يتم من خلال تقسيمه إلى مكونات أو وحدات منفصلة، ثم تنظيمها بطريقة متسلسلة، بحيث تعمل الواحدة تلو الأخرى في الدماغ البشري.

ومما يؤكد هذا الزعم أن تحليل النص لغويًّا يتطلب مرتكزًا أساسيًّا وهو تفكيك الوحدة النصية إلى مكوناتها الجزئية، بحيث تتيح لنا معرفة بنياتها الداخلية، وإن توخّي التحليل النحوي يتصل بتحليل الجملة إلى عناصرها الأولية للكشف عن علائقها وربط مكوناتها.

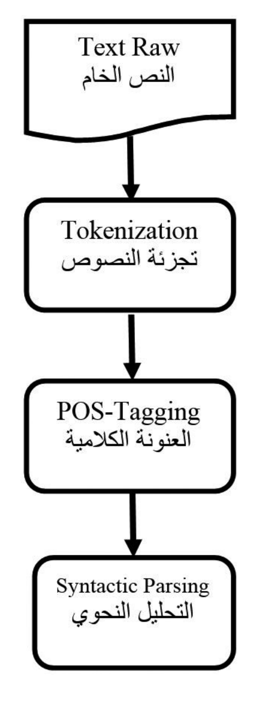
وفي إطار محاكاة الذكاء الاصطناعي للمخ البشري في تحليله للغة، وإدراكه لعناصر التركيب اللغوي، استمدت أساليب التعلم الآلي فكرتها من إدراك المخ البشري للنص اللغوي، التي تنطلق من تجزئة السلسلة النصية Text string إلى وحدات لغوية، ثم معالجتها في خطوات تتابعية؛ لنمذجة الاستعمال الإنساني للغة.

ويشير مفهوم تجزئة النصوص إلى تقسيم النص-آليا- إلى وحدات منفصلة من

خلال جملة من المعطيات اللغوية اللازمة التي تكون دليلا مستأنسًا للحاسوب في تعيينه إلى هذه الوحدات.(١)

وتنطلق معظم تطبيقات معالجة اللغات الطبيعية، مثل معنونات الأقسام الكلامية POS–taggers والمحللات النحوية Syntactic Parsers والمُجذِّعات Stemmers من النصّ المقسّم إلى أجزاء/ عناصر. وتشمل هذه الأجزاء الكلمات والأرقام وعلامات الترقيم والرموز، وغيرها من الوحدات المكوَّنة للنص.(٢)

إذا كانت تجزئة النصوص هي المطلب الرئيسي لتطبيقات معالجة اللغة الطبيعية، فإن دقة هذه التجزئة تنعكس على أداء التطبيقات اللغوية. ويوضح الشكل رقم ٦ مدى انعكاس أداء تجزئة النصوص على أداء التحليل النحوي.



الشكل:٦ خطوات عملية التحليل النحوي.

(1) Attia, M. (2007). Arabic Tokenization Systems. In proceeding of ACL. P.65. & Soudi, A. Bosch, A. Neumann, G¨. (2007). Arabic Computational Morphology Knowledge-based and Empirical Methods. Springer. P.32.

(2) Lüdeling, A. & Kytö, A. (2008). Ibid.P.527.

٢. مستويات تجزئة نصوص المدونة اللغوية عينة الدراسة

إن الهدف المنشود من هذه الدراسة هو التحليل النحوي للعربية الفصحى المعاصرة المرتكز على المفاهيم النحوية التي تعدُّ منطلقًا أساسيًّا للفهم الأتوماتي للنص اللغوي؛ لذلك فإن تجزئة النصوص قد اشتملت على ثلاثة مستويات أساسية:

٢-١ التجزئة على مستوى الجملة

«تعد الجملة من الناحية النحوية الوحدة المحورية لبنية النص». [١] غير أن مفهوم الجملة ليس واضحًا إلى الآن، فقد اضطربت تعريفات الجملة لدى النحاة اضطرابًا شديدًا، واتسعت مذاهبهم، تارة لاعتمادهم على أساس الناحية اللفظية المتمثلة في فكرة العامل، وأخرى لانطلاقهم من الناحية الدلالية. وفي ضوء هذه المنطلقات لتعريف الجملة يتضح الاختلاف في تحديد أبعاد الجملة.[٢]

وتتصل تجزئة الجملة بشكل عام بالتعرّف على معيار تحديد أبعاد الجملة الذي يمثله مقياس الشكل النحوي أو المعنى التام. ويتم التقسيم - في تحديد أبعاد الجملة في المدونة اللغوية عينة الدراسة- حسب الإسناد والتركيب التام المفيد، وما بين الجمل من علاقات الربط بواسطة أدوات الاستئناف والعطف.

ثم كان الانطلاق من علامات الترقيم لاسيما علامة النقطة، والاستفهام، والتعجب خير سبيل لتحديد هذه الأبعاد الجملية -تحديدا شكليًّا- في نصوص المدونة اللغوية، إذ تتخذ آلية تجزئة النصوص Tokenizer من علامات الترقيم وسيلة لتجزئة النص إلى جمل منفصلة.[٣] إلا أنَّ هذه الآلية تحتاج إلى تتمة الأدلة ليكتمل فيها من المحددات التي تبلغ حد الكفاية

(١) برينكر (كلاوس): التحليل اللغوي للنص، مدخل إلى المفاهيم الأساسية والمناهج، ترجمة: أ.د سعيد حسن البحيري، مؤسسة المختار للنشر والتوزيع، ٢٠١٠، ص٤٠.

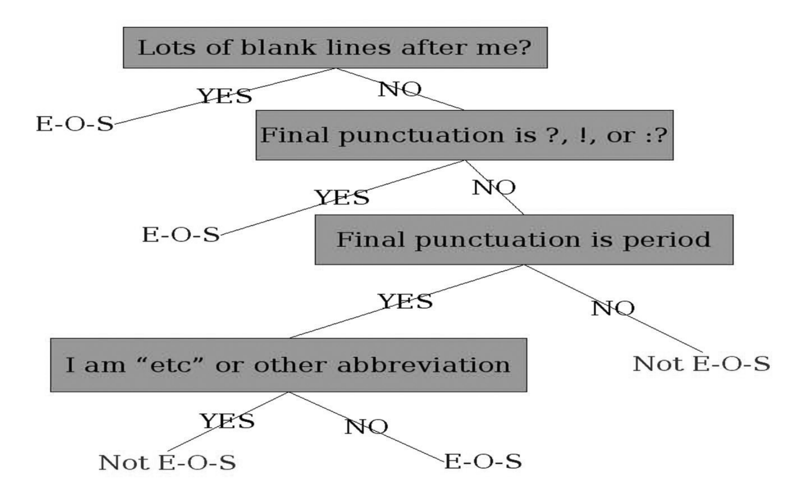
(٢) تعددت اتجاهات تعريفات الجملة، حيث أقام النحاة درسهم للجملة العربية على أساس الناحية اللفظية المتمثلة في الإعراب وفكرة العامل، فقسموا الأبواب النحوية وفق هذه الناحية إلى مرفوعات ومنصوبات ومجرورات ومجزومات، وقد دفعهم هذا المنهج إلى الانطلاق في درسهم من المبنى؛ ليصلوا إلى المعنى إلى أن جاء عبد القاهر الجرجاني في كتابه دلائل الإعجاز بنظرية تركيبية تنطلق من المعنى وصولا للمبنى أي باتجاه يتفق مع عملية الاتصال اللغوي.ينظر: حميدة (مصطفى): نظام الارتباط والربط في تركيب الجملة العربية، مكتبة لبنان-ناشرون، ١٩٩٧، ص٢. حول أبعاد الجملة عند النحاة القدماء. ينظر: عاشور (المنصف): السابق، ص٢١-٢٥. ينظر: عبادة (محمد إبراهيم). الجملة العربية، مكوناتها- أنواعها-تحليلها، مكتبة الآداب القاهرة، ط ٤، ٢٠٠٧م، ص ٢٩-٣٢.

(3) Habash, N. Faraj, R. & Roth, R. (2009). Syntactic Annotation in the Columbia Arabic Treebank. In Proceeding of ELDA.P.125.

في تجزئة النص إلى جمل؛ لما في علامات الترقيم من لبس يشوبها، فقد تعددت وظائفها في النص الكتابي بين وظيفتها الأساسية وما يتفرع منها، فعلى سبيل المثال لا الحصر النقطة التي تستخدم كمحدد للدلالة على نهاية الجملة، تستخدم بين الاختصارات مثل أ.د، ص.د.ب، وغيرها، وكذلك الفاصلة التي تعد ملمحا مميزا للفصل بين الوحدات أوالمكونات في الجملة، تستخدم حال الأرقام العشرية.

ويدفع الباحثون اللبس الناتج عن تلك الأطر باللجوء إلى أساليب الذكاء الاصطناعي؛ لقدرتها على استدعاء الاحتمالات لتلك الأطر، ومواجهتها للتعامل مع حالات اللبس المتعددة.

ومن تلك الأساليب الإحصائية لمواجهة الاحتمالات المتعددة في هذا الإطار المصنفات الثنائية binary classifiers التي تندرج تحتها تقنية مشجرات اتخاذ القرار[(١)] Decision Trees، حيث تعتمد تلك التقنية على تحديد خواص العلامات، وإعادة ترتيبها في صورة شجرية متدرجة، ويشكل التنسيق بين هذه الخواص تصورًا يؤدي إلى تحديد الفئات من حيث التجزئة أو عدم التجزئة[(٢)]. ويوضح الشكل رقم ٧ صورة توضيحية (شجرية) لتحديد خواص العلامات، ومن ثم تحديد تصنيف حالة الجملة من حيث نهايتها أو عدم نهايتها.



الشكل:٧ تحديد الكلمة النهائية في الجملة-من خلال الخواص- باستخدام مشجرات اتخاذ القرار.[(٣)]

(1) Nugues, M. (2010). An Introduction to Language Processing with Perl and Prolog: An Outline of Theories, Implementation, and Application with Special Consideration of English, French, and German (Cognitive Technologies).Springer.P.82.

(2) Nugues, M. (2010).Ibid.P.82.

ينظر: السعيد (المعتز بالله): مدونة معجم تاريخي للغة العربية « معالجة لغوية حاسوبية»، ص ١١٤.

(٣) مقتبس هذا الشكل من محتوى مادة معالجة اللغة الطبيعية التي يقدمها كلا من Dan Jurafsky & Chris Manning بجامعة ستانفورد.

حيث تعني E-O-S = End Of Sentence قرار نهاية الجملة الذي تؤسسه خوارزمية مشجرات اتخاذ القرار بناءً على خواص علامات الترقيم في الجملة من حيث كونها علاماتٍ للدلالة على انتهاء الجملة أو علامات لتكوين الاختصارات.

وقد استعانت الدراسة بآلية التجزئة المدرجة في محلل ستانفورد النحوي Stanford parser[١] التابع لفريق معالجة اللغات الطبيعية بجامعة ستانفورد؛ لتجزئة نصوص المدونة اللغوية إلى جمل منفصلة. إلا أنها لا تدعّم حالات اللبس الناجمة عن تلك الأطر المتمثلة في علامات الترقيم. وقد أمكن التعامل مع هذه الحالات -التي سبق ذكرها- في المدونة اللغوية بالمعالجة اليدوية.

٢-٢ التجزئة على مستوى الوحدات/ العناصر الرئيسية

العنصر اللغوي Token هو أصغر وحدة نحوية، يمكن أن تكون كلمة أو جزءًا من الكلمة، أو تعبيرًا اصطلاحيًّا، أو مركبًا، أو رمزًا، [٢] ومادامت العناصر اللغوية الرئيسية هي الجزء الملموس من التحليل فيمكن أن نطلق عليها أيضا "وحدات التحليل النحوي".[٣]

والوحدة الرئيسة هي البناء اللغوي المتكامل سواء أكانت كلمة أو علامة أو رقمًا، وتعد عنصرًا أساسيا في النص اللغوي. و تشتمل التجزئة على مستوى الوحدات أو العناصر الرئيسية Main Tokens على ثلاثة مستويات:

٢-٢-١ الكلمة

تعرَّف الكلمة[٤] في اصطلاح اللغويين بأنها" صيغة ذات وظيفة لغوية معينة في

---

(١) يمكن تحميله من خلال هذا الرابط:
Stanford-parser. (2015). http://nlp.stanford.edu/software/lex-parser.shtml

(2) Attia, M. (2007). Ibid.P.65.

(٣) شمس الدين(جلال): الأنماط الشكلية لكلام العرب، نظرية وتطبيقا دراسة بنيوية، مؤسسة الثقافة الجامعية، الأسكندرية، ط١، ١٩٩٥م، ص٦٨.

(٤) وقد عرفها أيضًا العالم الأمريكي»بلومفيلد« Bloomfield بأنها أصغر وحدة حرة، ومعنى هذا أن الكلمة عنده هي أصغر وحدة لغوية يمكن النطق بها معزولة، كما يمكن استعمالها لتركيب الجملة أو الكلام، ويجب أن تتكون من مورفيم حر Free Morpheme على الأقل. ينظر:
خليل (حلمي): الكلمة دراسة لغوية معجمية، دار المعرفة الجامعية، الأسكندرية، ١٩٩٨م، ص٦٨.

تركيب الجملة، تقوم بدور وحدة من وحدات المعجم، وتصلح لأن تفرد، أو تحذف، أو تحشى، أو يغير موضعها، أو يستبدل بها غيرها في السياق، وترجع مادتها غالبا إلى أصول ثلاثة".[١]

أمّا في عرف الحاسوب، فهي حيّز من الحروف المتشابكة، أو الحروف المفردة أو العلامات، أو الرموز، يحيطه من جانبيه مساحات بيضاء White Spaces. وهذه المساحات هي المعطيات التي تفضي إلى حدود الكلمة لتجزئة الوحدات الرئيسية في النص.[٢]

وثمة عديد من آليات تجزئة النصوص إلى وحدات رئيسية – وغالبًا هذه الآليات يتم إدراجها في المحللات النحوية – منها:

٢-٢-١-١ أداة التجزئة العربية Arabic Tokenizer .

عمل فريق معالجة اللغات الطبيعية بجامعة ستانفورد على تطوير هذه الأداة – المحرَّرة بلغة الجافا البرمجية Java[٣]– باعتبارها جزءًا أساسيًّا من المحلل التركيبي الذي يعمل على تجزئة الوحدات الرئيسية للنص المدخل، ومع ذلك يمكن استخدامها منفردة. وقد دعمت هذه الأداة الحروف العربية المشفرة بـصيغة UTF-8، كما تدعم العديد من بيئات التشغيل المتعددة. ويوضح الشكل رقم ٨ صورة لواجهة أسطر الأوامر أثناء تنفيذ هذه الأداة على ملف نصي.

(١) حسان(تمام): مناهج البحث في اللغة، مكتبة الأنجلو القاهرة، ط١، ١٩٥٥م، ص ٢٦٢.

(2) Jurafsky, D. & Martin, H. (2007). Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition. Prentice Hall .P.26. & Habash, N. & Faraj, R. & Roth, R. (2009). Ibid. P. 125. & Warner, C. & Lanfranchi, A. & O'Gorman, T. & Howard, A. & Gould, K. & Regan, M. (2002). Bracketing Biomedical Text: An Addendum to Penn Treebank II Guidelines. Institute of Cognitive Science, University of Colorado at Boulder. P. 3.

ينظر أيضا: السعيد (المعتز بالله): مدونة معجم تاريخي للغة العربية « معالجة لغوية حاسوبية»، ص ٨٤.

(٣) لغة الجافا Java هي لغة برمجية، قد تم تطويرها من قبل جيمس جوسلينج في مختبرات شركة صن ميكروسيستمز عام ١٩٩٢م، امتدادًا للغة السي بلس بلس ++C؛ لاستخدامها في بناء الأجهزة التطبيقية الذكية مثل التيلفزيون التفاعلي، الهواتف الذكية. ينظر:

Wikipedia The Free Encyclopedia. (2015). https://goo.gl/6gYUYi

```
ahmed@ahmed-PC: ~/Desktop/stanford-parser-full-2013-06-20
ahmed@ahmed-PC:~$ cd /home/ahmed/Desktop/stanford-parser-full-2013-06-20
ahmed@ahmed-PC:~/Desktop/stanford-parser-full-2013-06-20$ export CLASSPATH=./sta
nford-parser.jar
ahmed@ahmed-PC:~/Desktop/stanford-parser-full-2013-06-20$ java edu.stanford.nlp.
international.arabic.process.ArabicTokenizer <./Ahmed.txt >./foo.txt
Untokenizable:
Done! Tokenized 2 lines (134 tokens)
ahmed@ahmed-PC:~/Desktop/stanford-parser-full-2013-06-20$
```

الشكل:٨ أمر تنفيذ أداة التجزئة العربية Arabic Tokenizer



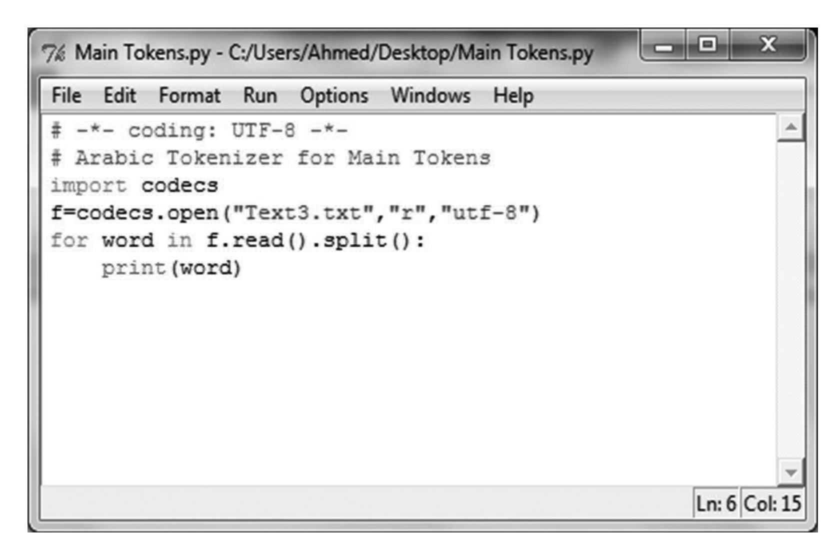
الشكل:٩ مخرجات أداة التجزئة العربية Arabic Tokenizer [١]

(١) يمكن تحميله من خلال هذا الرابط:
Arabic Tokenizer. (2015). http://nlp.stanford.edu/software/lex-parser.shtml

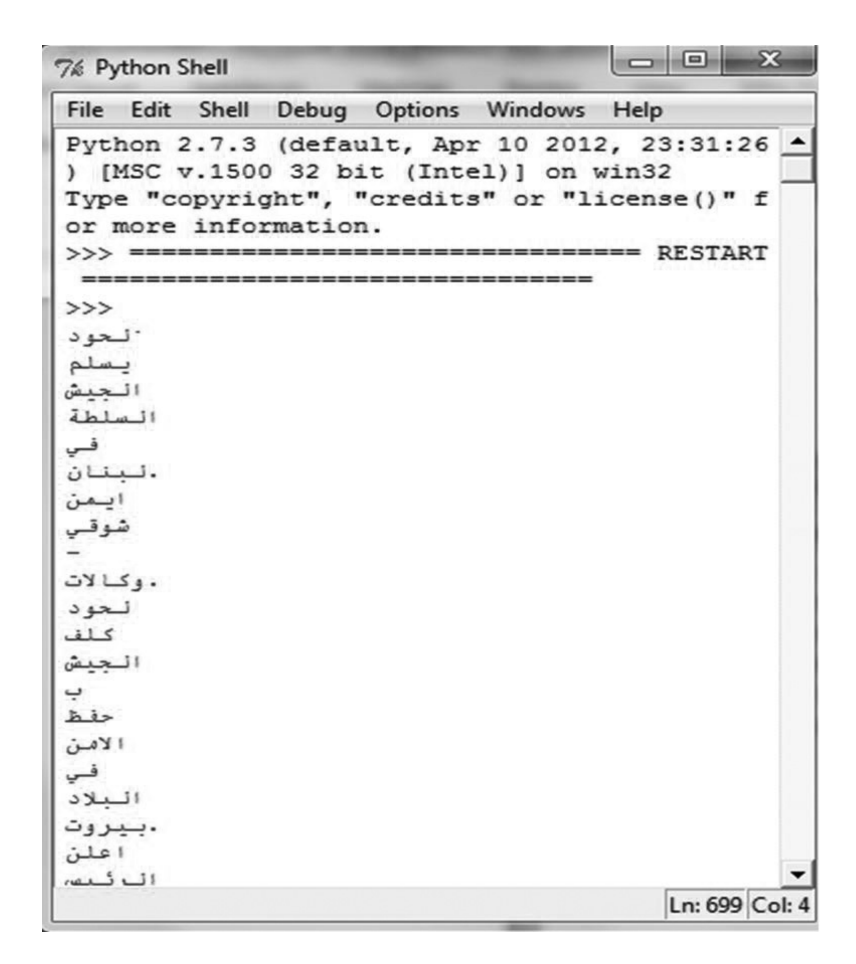
يظهر في الشاشة الأولى أو طرفية واجهة أسطر الأوامر Command lines أمرُ تنفيذ أداة التجزئة العربية المتضمن عنوان الملف النصي المراد تجزئته، وكذلك نتيجة التجزئة أو عدد الوحدات الرئيسية التي تم تجزئتها، أما في الشاشة الثانية فتُظهر الملف النصيّ بعد تجزئته إلى وحدات رئيسية.

٢-٢-١-٢ آلية تجزئة الوحدات الرئيسية Main Tokens.py

عمل الباحث على تطوير هذه الآلية - باستخدام لغة البايثون البرمجية Python - لتجزئة الوحدات الرئيسية في النص، مدعمة الحروف العربية المشفرة بـ UTF-8، كما تدعم العديد من بيئات التشغيل المتعددة. ويعرض الشكل رقم ١٠ أسطر الأداة البرمجية " الكود" داخل واجهة بايثون IDLE Python لكتابة أسطر الأوامر.



الشكل: ١٠ واجهة بايثون لكتابة أسطر الأوامر أسطرُ - كود - أداة تجزئة الوحدات الرئيسية Main Tokens.py



الشكل: ١١ مخرجات آلية تجزئة الوحدات الرئيسية Main Tokens.py

يظهر في الشاشة الأولي واجهة بايثون لكتابة أسطر الأوامر أسطرُ "كود" أداة تجزئة الوحدات الرئيسية، متضمنًا مكان الملف النصي المراد تجزئته، أما الشاشة الثانية فتُظهر الملف النصيّ بعد تجزئته إلى وحدات رئيسية.

٢-٢-٢ المركب غير الكلامي

"هو انضمام كلمة إلى كلمة فأكثر، وتكون بحكم المفرد نحويا ودلاليًّا"(١) مثل: عبد الله، جاب الله، أبو عيد، إسلام أون لاين، الصهيو أمريكي، الجيو إستراتيجية.

(١) الدحداح (أنطوان): معجم لغة النحو العربي، مكتبة لبنان ناشرون، ٢٠٠، ص ٢٩٤ (بتصرُّف).

فما الذي يبعث إلى انتهاج المركب غير الكلامي في المنهج المقترح للتجزئة؟ إنما يتقرر ذلك في ضوء الدراسة النحوية للجملة التي تستدعي بنية متناسقة من الوحدات النحوية الدالة على الوظيفية التي تحققها في بنية الجملة.

وتجدر الإشارة إلى أن تجزئة المركب غير الكلامي في نصوص المدونة اللغوية قد عُولجت بالمعالجة اليدوية من خلال وضع علامة الشرطة (-) بين الكلمة الأولى والكلمة الثانية؛ ليكونا في حكم الكلمة الواحدة. ويوضح الجدول رقم ٣ نماذج من تجزئة المركب غير الكلامي في نصوص المدونة اللغوية:

ما قبل المعالجة اليدوية	ما بعد المعالجة اليدوية
عبد الله	عبد-الله
السات كوم	السات-كوم
عبد الرحمن	عبد-الرحمن
إسلام أون لاين	إسلام-أونلاين
غلام الله	غلام-الله
الصهيو أمريكي	الصهيو-أمريكي
شمس الدين	شمس-الدين
عطا الله	عطا-الله
عبد المحسن	عبد-المحسن
حزب الله	حزب-الله
أبو تريكة	أبو-تريكة
الجيو إستراتيجية	الجيو-إستراتيجية

الجدول :٣ معالجة المركب غير الكلامي في نصوص المدونة اللغوية

٢-٢-٣ الرمز أو العلامة:

يشمل جميع الرموز المستخدمة في النص العربي، مثل علامات الترقيم والأرقام، وغيرها من الرموز.

ووفقًا لاعتبار البنية النحوية وحدةً للتحليل، استوجب على الدراسة توحيد نسق الأرقام والتواريخ في نصوص المدونة اللغوية، بحيث تمتلك حيّزًا كتابيًّا واحدًا على اعتبار أنها وحدة نحوية مفردة، وقد أفضى هذا التوحيد إلى إزالة الفراغات البيضاء بين الأرقام مع توحيد أشكالها. ويوضح الجدول رقم ٤ نماذج من اتساق الأرقام المتعددة والتواريخ في نصوص المدونة اللغوية.

ما قبل المعالجة اليدوية	ما بعد المعالجة اليدوية
2, 4	2,4
2008 -1 -21	21-1-2008
/1 / 2008 21	21-1-2008
% 30	30%
50 - 50	50-50
4 / 19	19/4
% 53 - 46	46-53%
0 .11	0.11
12 , 33	12 ,33

الجدول:٤ معالجة الرمز أو العلامة في نصوص المدونة اللغوية

٢-٣ التجزئة على مستوى الوحدات/ العناصر الفرعية

يمكن أن نعرّف العنصر اللغوي أيضا بأنه" بناء لغوي يحدده مستوى التحليل"[١]، إذ نجد أن العنصر اللغوي الرئيسي قد يكون مكونًا من مورفيم/ عنصر فرعي واحد أو أكثر من مورفيم، فعلى سبيل المثال يمكن للكلمة المفردة (العنصر الرئيسي) أن تشمل أربع وحدات فرعية سواء أكانت سوابق أو لواحق. [٢]

وتتوقف حدود عملية تجزئة العناصر الرئيسية إلى عناصر فرعية على طبيعة الغرض من البحث، أي ما العناصر الفرعية المراد تجزئتها من العناصر الرئيسية؟ ويقتضي لتحليل الجملة العربية تجزئة عناصرها الأساسية التي تكوّن العلاقات النحوية في بنية الجملة.

ولما كان الكلام سلسلة من الجزئيات المتتابعة، كان لزامًا على تلك الدراسة أن تعرض أنوع تلك الجزئيات:

ثمة أنواع من المورفيمات اللصقية Concatenative Morphemes في اللغة العربية: الجذع((Stem واللواصق((affixes والزوائد (Clitics): [٣]

٢-٣-١ الجذع Stem : هو جزء أساسي من الكلمة، يأتي مشتقًا أو جامدًا، وينتج عن اتحاد المورفيمات اللصقية للكلمة، ومن أمثلته: الجذع (كتب) الذي تكون عنه التركيب في(وسيكتبونها) والجذع (مكتب) في صيغة الجمع (المكتبات).

٢-٣-٢ اللواصق Affixes: هي مورفيمات تتعلق بجذع الكلمة، وهناك نوعان من اللواصق[٤]:

١)**السوابق(Prefixes)**: والسابقة مورفيم يسبق الجذع في أوله، ومن أمثلته: نون في الفعل المضارع في "نفعل- نعمل- نشكر".

٢)**اللواحق(Suffixes)**: واللاحقة مورفيم يلحق الجذع في آخره، ومن أمثلته: الواو والنون في جمع المذكر السالم في" المسلمون-العاملون".

---

(١) شمس الدين(جلال): السابق، ص ٦٩.

(2) Attia, M. (2007). Ibid. P.65.

(3) Habash , N. (2010). Ibid.P.41. & Jurafsky, D. & Martin, H. (2007). Ibid.7.

(٤) هناك نوع ثالث من اللواصق، لكنه يتوسط الجذع، ولا يتعلق به مثل ألف اسم الفاعل مثل كاتب، وهو الدواخل (Infixes).

٢-٣-٣ الزوائد Clitics: هي مورفيمات نحوية تكون مقيدة بكلمات أخرى، وتتعلق بجذع الكلمة بعد اللواصق. وهناك نوعان من الزوائد[١]:

١)**الزوائد في بداية الكلمة (Proclitics)** فهي تشبه اللواصق، ولكنها تختلف اختلافًا واضحا عن اللواصق التي تمثل جزءًا من الكلمة صوتيًّا وبنيويًّا، ومن أمثلتها: حروف العطف، وحروف الجر، والنداء.

٢)**الزوائد في نهاية الكلمة (Enclitics)** وهي التي تعقب الكلمة، مثل الضمائر المتصلة.

وقد يكون القرار مربكا أحيانا في جعل المورفيم لاصقة صرفية أو زائدة نحوية، ومع ذلك نستطيع أن نقول – عموما – أن اللواصق تحمل ملامح صرفية نحوية مثل (الزمن – الشخص – الجنس – العدد)، بينما الزوائد تخدم الوظائف النحوية مثل (النفي – التعريف – العطف أو الجر).[٢]

٢-٣-١ المنهجيات المختلفة لتجزئة الزوائد Decliticization في النص

تجري الزوائد النحوية Clitics في بناء الجملة العربية مجرى الأساس في البناء المتكامل، إذ هي ضرب من الملفوظات المكونة لأبنية الجملة، وتكون مرهونة بالوظيفة التركيبية في تواليها المختلفة.

وممَّا هو جدير بالذكر أن تعدد منهجيات تجزئة الزوائد النحوية Clitics يخضع لطبيعة الهدف المأمول من تلك التجزئة، "فما هو مناسب لنظم استرجاع المعلومات ربَّما لا يتناسب مع نظم الترجمة الآلية الإحصائية، وهكذا"[٣]. ويعرض الجدول رقم ٥ بعضًا من منهجيات التجزئة المستخدمة في معالجة اللغة العربية آليًّا بشكل عام.

(1) Attia, M. (2007). Ibid. P.65.

(2) Attia, M. (2007). Ibid. P.65.

(3) Olive, J. & Christianson, C. & McCary, J. (2011). Handbook of Natural Language Processing and Machine Translation. Springer. P.81. And Habash , N. (2010). Ibid.P.78.

تجزئة الزوائد Clitics							المنهج
الضمائر	همزة الاستفهام	أل التعريف	السين التسويفية	حروف الجر	حروف العطف		الزوائد
					الواو	الفاء	
×	√	×	×	×	√	√	D1
√	√	×	√	√	√	√	D2
√	√	√	√	√	√	√	D3/S1
√	√	√	√	√	√	√	S2
×	√	×	×	×	√	×	WA
√	√	×	×	√	√	√	ATB old
√	√	×	√	√	√	√	ATB

الجدول: ٥ المنهجيات المختلفة لتجزئة الزوائد النحوية.

حيث تعني Declitization degree 3,2,1 [1] (D1,D2,D3) بتجزئة الزوائد ذي الدرجات المختلفة، فمنهجية الدرجة الأولى D1 تقوم على تجزئة (حروف العطف، وهمزة الاستفهام) من الزوائد النحوية، أما منهجية الدرجة الثانية D2 فتقوم على تجزئة (حروف العطف، وحروف الجر الملتصقة بالكلمة، والسين التسويفية، وهمزة الاستفهام، والضمائر)، بينما منهجية الدرجة الثالثة D3 تقوم على فصل الزوائد النحوية كلها.

(1) Olive, J. & Christianson, C. & McCary, J. (2011). Ibid. P.81.

أمَّا (S1, S2,)[١] Segmentation degree 1,2 تعنيان بتقسيم الزوائد ذي الدرجات المختلفة، فمنهجية الدرجة الأولى والثانية S1 S2 تتشابه مع منهجية D3، إلا أن S2 تجعل زوائد بداية الكلمة في متوالية واحدة، مثال ذلك تجزئة كلمة "وسينتهي" لتصبح " وس+ ينتهي ".ومنهجية (WA) [٢] Decliticizing the conjunction wa تعني بتجزئة حرف العطف "الواو"، وهمزة الاستفهام، وتتشابه مع منهجية D1 إلا أن الأخيرة تفصل حرفي العطف الفاء والواو.

أمَّا ATB old (Arabic Treebank old) و ATB[٣] (Arabic TreeBank) فتعنيان بتجزئة (حروف العطف، وحروف الجر الملتصقة بالكلمة، وهمزة الاستفهام، والضمائر) إلا أن ATB تزيد عنها زائدة السين التسويفية.

ونخلص إلى أن هناك اتفاقًا بين منهجي D3 / S1 و S2 وكذلك بين D2 و ATB في تجزئة الزوائد بنوعيها، ويتدرّج الاتفاق إلى نوع من الاختلاف بين D1 و WA ، إذ يتمثل نمط الاختلاف بينهم باقتصار WA على الفاء دون الواو في حروف العطف، وكذلك بين ATB old و ATB، حيث يتميّز ATB بتجزئة السين التسويفية.

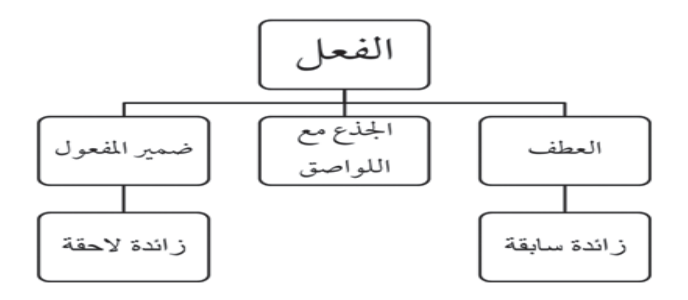
وطبيعة ما نصبو إليه – وهو التحليل النحوي للجملة – يجعلنا نقف أمام الوحدات النحوية (الزائدة بنوعيها) باعتبارها عنصرًا مستقلًّا للتحليل النحوي ذا علاقات نحوية نظمية بغيرها. وقد انتهجت الدراسة نهج البنك الشجري العربي من جامعة بنسلفانيا في تجزئة الزوائد، والتي تمثلت في (أداة التعريف–همزة الاستفهام – حروف الجر– حروف العطف – الضمائر)، لكنَّ الدراسة لم تتطرق لفصل أداة التعريف؛ نظرًا لكونها وحدة لسيت دالة على وظيفة نحوية في بنية الجملة.

وفي ضوء المنهج المقترح للتجزئة، يمكن أن تصل الوحدات الفرعية الممكنة في الفعل إلى ثلاث وحدات فرعية (مثال ذلك: وسنقولها)، كما في الشكل رقم ١٢:

(1) Badr, I. & Zbib, R. & Glass, J. (2008). Segmentation for English-to- Arabic Statistical Machine Translation. In Proceedings of ACL. P.154.

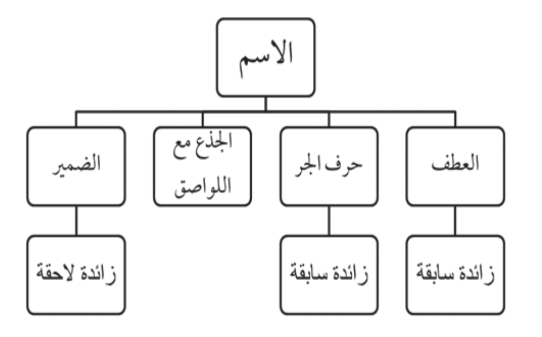
(2) Olive, J. & Christianson, C. & McCary, J. (2011). Ibid.P.81.

(3) Maamouri, M. & Bies, A. & Buckwalter, T. & Mekki, W. (2004). The Penn Arabic Treebank: Building a Large-Scale Annotated Arabic Corpus.



الشكل:١٢ الوحدات الفرعية الممكنة للفعل.

وكذلك يمكن أن تصل الوحدات الفرعية الممكنة في الاسم إلى أربع وحدات فرعية (مثال ذلك: وبمسؤوليتهم)، كما في الشكل رقم ١٣:



الشكل:١٣ الوحدات الفرعية الممكنة للاسم.

٣. التوحيد الهجائي Orthographic Normalization

تتميز الكتابة العربية بحساسية سياقية شديدة لكونها تمتاز بصفة التشبيك، إذ يتغير شكل الحرف الواحد تبعا للحرف السابق له، أو اللاحق له، أو كليهما، وهذا مما يتتطلب تعدد أشكال الحروف.[١]

وتفقد بعض الكلمات شكلًا من حروفها حال انضمامها إلى مورفيم آخر أو تجزئتها من ذاك المورفيم. ولذا يلجأ الباحثون في معالجة اللغة الطبيعية دائمًا إلى التوحيد الهجائي بإرجاع تلك الكلمات إلى شكلها الصحيح بعد التجزئة، بهدف الحد من الضوضاء

(١) علي (نبيل): اللغة العربية والحاسوب، ص٢٠٤. (بتصرُّف).

وتناثر البيانات في النصوص. [١]

وتتمثل أنماطها فيما يلي:

١- حرف الجر الملتصق بأل التعريف في " للعمل" يتم تجزئتها إلى " ل+لعمل" وبعد التوحيد الهجائي تصبح " ل+العمل".

٢- حالات الهمزة المتطرفة الملتصق بها ضميرحسب موقعها في الإعراب مثل" سماؤه" "سماءه" "سمائه" يتم تجزئتها إلى "سماؤ+هـ" "سماء+هـ" "سمائ+هـ" وبعد التوحيد الهجائي تصبح " سماء+هـ" .

٣- الألف المقصورة الملتصقة بالضمير في " مصطفاهم " يتم تجزئتها إلى " مصطفا+هم" وبعد التوحيد الهجائي تصبح " مصطفى+هم".

٤- التاء المربوطة الملتصقة بضمير في " مدرستنا" يتم تجزئتها إلى " مدرست+نا" وبعد التوحيد الهجائي تصبح " مدرسة+نا".

وقد استعانت الدراسة بآلية [٢] MADA+TOKAN لتجزئة الوحدات الفرعية في نصوص المدونة اللغوية - عينة الدراسة - التي تم تطويرها من قبَل فريق معالجة اللغات الطبيعية بمركز أنظمة التعلم الحاسوبي بجامعة كولومبيا CCLS، وقد حرّرت هذه الآلية بلغة بيرل البرمجية Perl[٣]، كما تدعم بيئة التشغيل يونكس فحسب. ويوضح الشكل رقم ١٤ صورة لواجهة أسطر الأوامرأو طرفية لينكس أثناء تنفيذ هذه الأداة على ملف نصي.

(1) El Koly, A. & Habash, N. (2010). Techniques for Arabic Morphological Detokenization and Orthographic Denormalization. In Proceeding of LREC Workshop on Semitic Languages.P.45-46. And Habash , N. (2010). Ibid.P.21.

(٢) شرع مركز أنظمة التعلم الحاسوبي بجامعة كولومبيا عام ٢٠٠٥م - بقيادة العلَمَيْنِ نزار حبش وأون رامبو Owen Rambow - في بناء نظام متعدد المهام (التحليل الصرفي، والتشكيل الآلي، وتجزئة النصوص، وفك اللبس التصريفي، وتعيين الأقسام الكلامية، والتجذيع، والتفريع) بحيث يكون متاح للأغراض البحثية، وقد حقق نتائج لا بأس بها على مستوى التشكيل الآلي. ثم ظهر في ثوب جديد عام ٢٠١٥، حيث أدمج في نظام AMIRA لأدوات معالجة النص العربي التابع لجامعة كولومبيا أيضًا؛ ليظهر في إصدار جديد باسم MADAMIRA محرّرًا بلغة الجافا البرمجية بدلًا من لغة بيرل. وقد دعم هذا الإصدار الجديد العربية الفصحى والعامية المصرية. ينظر:
MADAMIRA. (2015). http://nlp.ldeo.columbia.edu/madamira

(٣) لغة بيرل Perl هي لغة برمجية، قد تم تطويرها من قبل لاري وول عام ١٩٨٧م، بهدف معالجة النصوص في نظام يونكس Unix التي كان من الصعب - حينها - معالجتها بالوسائل المستعملة. ينظر:
Wikipedia The Free Encyclopedia. (2015). https://goo.gl/Ektxqk

```
ahmed@ahmed-PC: ~/Desktop/MADAIn/MADA-3.2
Done.
-- Finished SVM tagging.  Time: 63 wallclock secs ( 0.00 usr  0.00 sys + 58.78 cusr  1.06 csys = 59.84 CPU)
==========

Selecting Analysis...
Done.
-- Finished Analysis Selection.  Time:  0 wallclock secs ( 0.01 usr  0.01 sys +  0.78 cusr  0.06 csys =  0.86 CPU)
==========

-- Finished MADA. Total Time: 70 wallclock secs ( 6.02 usr  0.18 sys + 60.09 cusr  1.14 csys = 67.43 CPU)

==========

Running TOKAN

Parsing TOKAN scheme(s)...
# Input <token-def> =( SCHEME=ATB )
Loading database [/home/ahmed/Desktop/MADAIn/MADA-3.2/MADA/almor.db] in [generation] mode ...
...Finished loading ALMOR database
Reading MADA file to tokenize...

Finished TOKAN. Time:  6 wallclock secs ( 0.00 usr  0.00 sys +  6.71 cusr  0.12 csys =  6.83 CPU)

==========

Finished All of MADA+TOKAN.  Final Total Time: 77 wallclock secs ( 0.00 usr  0.00 sys + 73.02 cusr  1.48 csys = 74.50 CPU)
 All output files can be found under /home/ahmed/Desktop/MADAIn/MADA-3.2/Out

==========

ahmed@ahmed-PC:~/Desktop/MADAIn/MADA-3.2$
```

الشكل:١٤ طرفية واجهة أسطر الأوامر «لينكس» أسطر معالجة الأداة MADA+TOKAN للملف النصي(١)

الشكل:١٥ مخرجات برنامج MADA+TOKAN

(١) يمكن تحميلها من خلال هذا الرابط:
MADA+TOKAN. (2015). http://www1.cs.columbia.edu/~rambow/software-downloads/MADA_Distribution.html

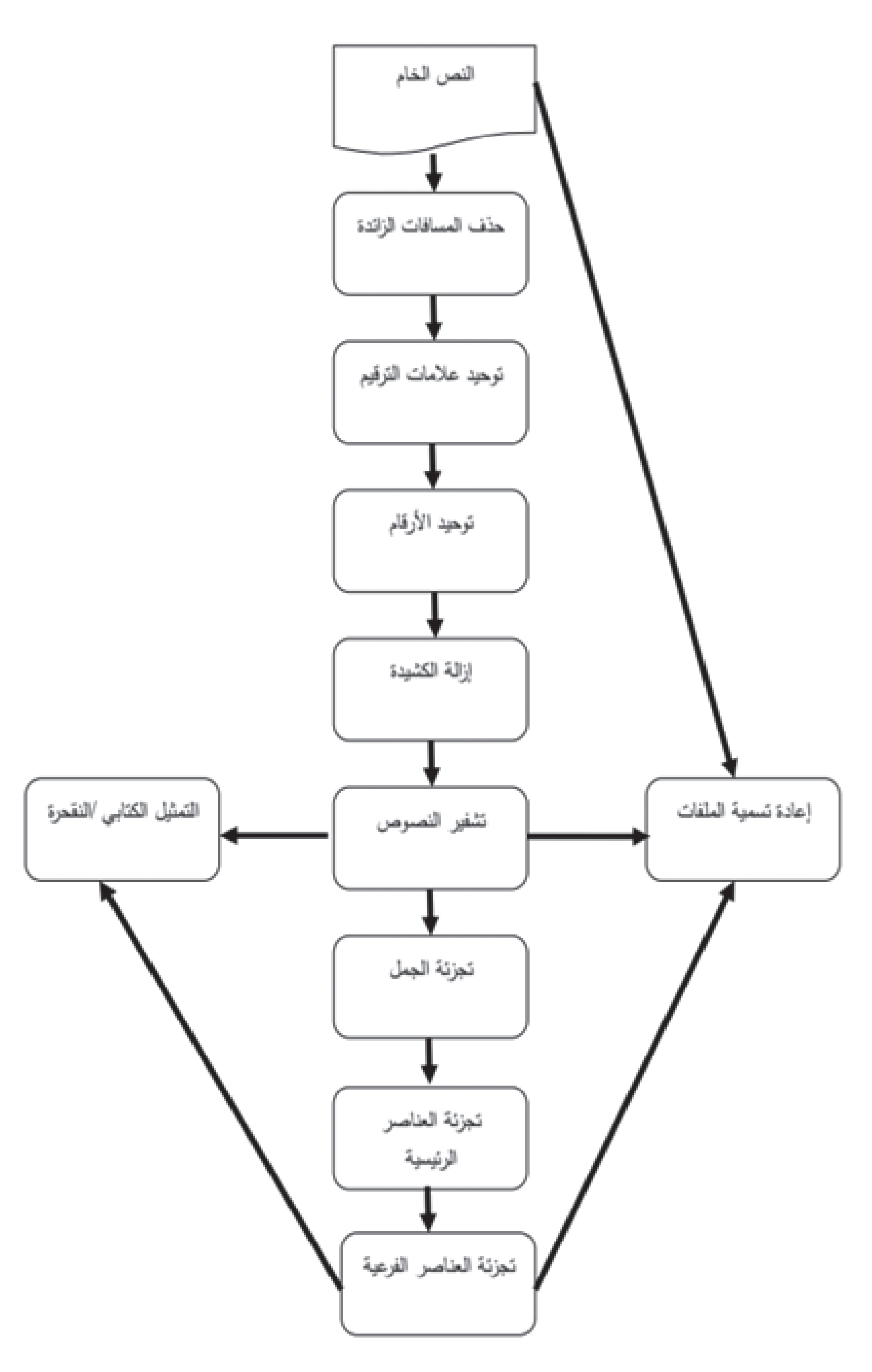
يظهر في الشاشة الأولى طرفية واجهة أسطر الأوامر"لينكس" أسطر معالجة الأداة للملف النصي الذي يمر بمرحلتين متتاليتين لتجزئته، المرحلة الأولي التحليل الصرفي Morphological analysis باعتباره المسؤول عن تحديد المعطيات الأساسية حول أجزاء الكلمة، ثم مرحلة التجزئة Running Token التي تنطلق من معطيات أجزاء الكلمة لا سيما اللواصق والزوائد. أما الشاشة الثانية فتُظهر الملف النصيّ بعد تجزئته للوحدات الفرعية.

وقد أمكن تعديل نسبة الخطأ الناتجة عن الآلية – بالمعالجة اليدوية في نصوص المدونة اللغوية – والتي تركزت بشكل كبير على الحالات التي يشوبها اللبس مثل الواو التي تأتي في أول الكلمة؛ لكنها جزء من جذع الكلمة " وَدَّعَت" وكذلك الباء في كلمة "بَسَنْت" وكلمة "بِطَاقة"، وكذلك الكاف في كلمة "كَفْر"، والهاء في كلمة "كره" وغيرها.

٤. نماذج لتجزئة الوحدات اللغوية في المدونة اللغوية عينة الدراسة

الجملة١	وكان الشيْخ جاب اللهّ قد أخرج عبر رفاقه من حركة " النهْضة ".
التجزئة١	و+كان الشيْخ جاب- اللهّ قَد أخرِج عبر رفاق+هـ من حركة " النهضة ".
الجملة٢	ورفضت السلطات السماح لقيادات وعناصر من " جبهة الإنقاذ الإسلامية " بالترشح في الانتخابات على قوائم أحزاب صغيرة.
التجزئة٢	و+رفضت السلطات السماح ل+قيادات و+عناصر من " جبهة الإنقاذ الإسلامية " بِ+الترشح في الانتخابات على قوائم أحزاب صغيرة.
الجملة٣	وهي شركات محْدودة التعامل تم شطبها لأسباب تتعلق بنقص شروط القيد.
التجزئة٣	و+هي شركات محْدودة التعامل تم شطب+ها ل+أسباب تتعلق ب+نقص شروط القيد.
الجملة٤	أهي بالملايين؟
التجزئة٤	أ+هي بِ+الملايين؟
الجملة٥	وذكرت صحيفة "ذا ديلي تلِيجراف" البريطانية يوم ٢١ – ١ – ٢٠٠٨ أن المملكة "تعتزم بنهاية العام ٢٠٠٨، رفع الحظر المفروض على قيادة النساء للسيارات.
التجزئة٥	و+ذكرت صحيفة "ذا- ديلِي- تلِيجراف" البرِيطانِية يوم ٢١-١-٢٠٠٨ أن المملكة "تعتزم ب+نهاية الْعام ٢٠٠٨، رفع الحظر المفروض على قيادة النساء ل+السيارات.

الجدول:٦ أمثلة على التجزئة من المدونة اللغوية عينة الدراسة.



الشكل:١٦ خريطة تدفق لمعالجة نصوص المدونة اللغوية

٥ . التحليل الإحصائي لوحدات Tokens المدونة اللغوية

”تتسم اللغة بخاصية الانتظام الإحصائي، حيث تظهر أي عينة أمينة في تمثيلها لواقع الاستخدام اللغوي استقرارًا إحصائيًّا من حيث معدلات تواتر الوحدات اللغوية المختلفة“[(١)]، سواء أكانت كلمات أو وحدات أو جمل.

وبعد أن استقرت المدونة اللغوية في صورتها الأخيرة أي بعد عملية تجزئة النصوص، يمكننا إحصاء وحدات مادة المدونة اللغوية بما يساعد على معرفة مدى التناسق بين مجالات المدونة اللغوية المختلفة. ويعرض الجدول رقم ٧ بيانات أو إحصاءات وحدات المادة اللغوية عينة الدراسة وفق التصنيف المجالي:

بيانات وحدات المدونة اللغوية[(٢)] -عينة الدراسة- وفق التصنيف المجالي.

م	المجال	عدد الكلمات	النسبة المئوية≅	عدد الوحدات	النسبة المئوية≅	عدد الجمل	النسبة المئوية≅
١	السياسة	50560	47.8%	59465	50.49%	2480	44.92%
٢	الاقتصاد	11597	10.96%	11882	10.09%	632	11.45%
٣	الفنون	11347	10.73%	10920	9.27%	635	11.50%
٤	الاجتماع	11147	10.54%	12746	10.82%	648	11.74%
٥	العلوم	11290	10.67%	11705	9.94%	592	10.72%
٦	الرياضة	9840	9.30%	11067	9.40%	534	9.67%
-	المجموع	105781	100.00%	117785	100.00%	5521	100.00%

الجدول:٧ إحصائيات وحدات المدونة اللغوية

(١) علي (نبيل)، حجازي (نادية): الفجوة الرقمية، رؤية عربية لمجتمع المعرفة، ص ٣١٧.

(٢) الكلمة هي البناء اللغوي المتكامل في بنية الجملة أي الصورة الأولى لشكل المفردات في المدونة اللغوية، أما الوحدة فهي أصغر وحدة نحوية يمكن أن تكون كلمة أو جزءًا من كلمة أو تعبيرا اصطلاحيا أي الصورة الثانية لشكل المفردة بعد التجزئة. وقد قمت بعمل إحصاء لشكل المدونة اللغوية قبل التجزئة كما ورد في جدول:١ وكذلك بعد التجزئة كما في جدول:٧.

يلاحظ أن أنصبة التوزيع النسبي لوحدات نصوص المدونة اللغوية (الرئيسية والفرعية)، تميل إلى الثبات أو الاستقرار في نصوص مجالات المدونة اللغوية، حيث نرى تقارب المنحنى بين الكلمات ووحدات نصوص المدونة، وكذلك تميل جمل المدونة إلى الانتظام عبر مجالات المدونة المختلفة من الناحية الإحصائية.

الفصل الثاني

العنونة بالأقسام الكلامية

مقدمة:

لما كان التحليل النحوي / التركيبي يتطلب تعيين الأقسام الكلامية لأجزاء الجملة باعتبارها مدخلًا ابتدائيًا لتحديد العلاقات التركيبية الناشئة فيما بينها، فقد جاء هذا الفصل معنونًا بـ (العنونة بالأقسام الكلامية) تتناول الدراسة فيه توضيح تقسيم الكلام وأهميته، وكذلك الحاجة إليه، ثم مفهوم العنونة بالأقسام الكلامية، وأهميتها في تطبيقات المعالجة الآلية للغة الطبيعية بشكل عام، ثم تعرض الفئات الكلامية المختلفة للغة العربية المستخدمة في العديد من التطبيقات، وأخيرًا يعرض الفصل نماذج من عنونة المدونة اللغوية بالعنونة الكلامية. وقد جاء هذا الفصل في ثلاثة مباحث على النحو التالي:

١- المبحث الأول: **تقسيم الكلام وأهميته.**

- ما الحاجة لتقسيم الكلام؟

٢- المبحث الثاني: **العنونة بالأقسام الكلامية.**

- فئة خوجة الكلامية.
- فئة باكولتر الكلامية.
- فئة بييز الكلامية.

- فئة بادت الكلامية.
- فئة آر دي آي الكلامية
- الفئة الكلامية المقلصة الموسعة.
- فئة القريني الكلامية.
- فئة كاتب الكلامية.
- فئة كاليك الكلامية.
- الفئة الكلامية للنص القرآني.
- فئة سلمة الكلامية.

٣- المبحث الثالث: **عنونة المدونة اللغوية - عينة الدراسة - بالأقسام الكلامية.**

المبحث الأول

تقسيم الكلام وأهميته

١ – تقسيم الكلام وأهميته

أفضى تفكير النُّحاة العرب في النَّظر لِلكلام ومجاريه النَّحوِيَّة إلى أنَّ الكلام[١] يأتلف من ألفاظ متنوعة متباينة في الخصائص والمعاني والوظائف[٢]، تربطها علاقات سياقية منطقية أو يتخلَّلها بعض الألفاظ لاصطناع تلك العلاقة، وأنّ المفردة هي الوحدة القاعدية التي تبنى عليها العلاقات التركيبية، فكان استقرارهم في اللُّغة على تمييز مكوناتها ووحداتها في أنظمة متكاملة؛ لتحرّي معطيات اللُّغة ونظمها.

ولما كان غرضهم من الاستقراء والوصف تعليميًّا فحسب، فقد أولوا اهتمامًا

(١) يلاحظ «أن النَّحاة العرب قد استعملوا الكلام بمعنى الكلمات أحيانًا، فسمّوا تقسيم الكلمات إلى اسم وفعل وحرف تقسيمًا للكلام»، حسان(تمام): مناهج البحث اللغوي، مكتبة الأنجلو المصرية، ١٩٩٠م، ص١٩٦.

(٢) يقول ابن السَّراج في هذا السياق «والذي يأتلف منه الكلام الثلاثة، الاسم والفعل والحرف، فالاسم قد يأتلف مع الاسم نحو قولك : «الله إلهنا» ويأتلف الاسم والفعل نحو: قام عمرو، ولا يأتلف الفعل مع الفعل، والحرف لا يأتلف مع الحرف، فقد بان فروق ما بينهما» ابن السراج (أبي بكر محمد بن سهل): الأصول في النحو، تحقيق د.عبد الحسين الفتلي، مؤسسة الرسالة، ط٣، ١٩٩٣م، الجزء الأول، ص٤١. ويذكر أيضا عبد القاهر الجرجاني في مدخله لدلائل الإعجاز أن «الكلام ثلاث: اسم، وفعل، وحرف. وللتعليق فيما بينهما طرق معلومة؛ وهو لا يعدو ثلاثة أقسام: تعلق اسم باسم، وتعلق اسم بفعل، وتعلق حرف بهما» ثم شرع في ذكر حالات التعلق بالاسم والفعل والحرف. ينظر: الجرجاني (عبد القاهر): دلائل الإعجاز، مطبعة المدني بالقاهرة، ١٩٩٢م، ص ٤-٨.

بالتَّجريد والتَّصنيف اللَّذين ينكشفان بالتَّدرُّج التَّتابعي في النَّظر لِلُّغة، ثم كان تقسيم الكلم Part-Of-Speech (POS) بمثابة المعطيات الأوليَّة للتَّحليل والوصف.

أجمع النَّحاة العرب القدماء على أن الكلام ينقسم إلى ثلاثة أقسام[١] : اسم – فعل –حرف، وتنوَّعت تعريفات تلك الأقسام حسب المقاييس التي استندوا إليها كالمعنى والشَّكل الصَّرفيِّ والنَّحويِّ، والإعراب والعمل النَّحويِّ، فيعرّف ابنُ هشام والسِّيوطيُّ وغيرهم من العلماء «الاسم بأنها كلمة تدل على معنى في نفسها ولا تقترن بزمان، وإذا اقترنت ففعل، أو في غيرها بأن احتاجت في إفادة معناها إلى اسم أو فعل أو جملة فحرف».[٢] بينما يعرّف ابن جني «الاسم ما حسن فيه حرف من حروف الجر أو كان عبارة عن الشخص... والفعل ما حَسُنَ فيه قد أو كان أمرًا، والحرف ما لم تحسن فيه علامات الأسماء ولا علامات الأفعال، وإنَّما جاء لمعنى في غيره نحو هل وبل وقد».[٣] ويتَّضح من خلال هَذَيْنِ التَّعريفَيْنِ أن النَّحاة العرب اعتمدوا في تقسيمهم على المعنى تارة والمبنى تارة أخرى، وهذا ينبئ عن عدم تفسيرهم للظَّواهر اللُّغوِيَّة الهامة في واقعها اللغوي؛ لأنَّهم انطلقوا في دَرْسِهم من المبنى وصولًا لِلْمَعْنى عكس ما تجري عليه عملية الاتِّصال اللغوي. كما اصطبغت جل هذه التَّعريفات بالمنطق الأرسطي، إذ أخذ النَّحو التَّقليديُّ الجملة الخبريَّة باعتبارها أساسَ البحث اللُّغويِّ، ومن ثمَّ تحدَّد أقسام الكلام حسب وظيفتها في هذه الجملة فقط.[٤]

ويخرج بعض اللُّغوِيِّين المُعاصِرين عن هذا الإطار التَّقسيميِّ، معتمدين على النَّهج المُتوازي بين الشَّكل والمعنى؛ لِيَصلُوا إلى تقسيم أدقّ من تقسيم النَّحاة القدماء، فيتَّجِهُ

(١) يعد سيبويه أول من أشار إلى التقسيم الثلاثي في العربية، إذ يقول في كتابه «فالكلم: اسم، فعل، حرف جاء بمعنى ليس باسم ولا فعل». سيبويه (أبو بشر عمرو بن عثمان بن قنبر): الكتاب، تحقيق عبد السلام محمد هارون، مكتبة الخانجي بالقاهرة، ١٩٨٨م، الجزء الأول، ص ١٢.

(٢) ينظر: السيوطي (الإمام جلال الدين): همع الهوامع في شرح جمع الجوامع، تحقيق د.عبد العال سالم مكرم، مؤسسة الرسالة، ١٩٩٢م، الجزء الأول، ص٣.

(٣) ابن جني (أبو الفتح عثمان): اللمع في العربية، تحقيق د.سمير أبو مغلي، دار مجدلاوي للنشر، ١٩٨٨م، ص ١٥-١٦.

(٤) الراجحي (عبده): النحو العربي والدرس الحديث «بحث في المنهج»، دار النهضة العربية للطباعة والنشر، ١٩٧٩م، ص٤٦-٤٧.

الدُّكتور إبراهيم أنيس إلى تقسيمٍ رباعيٍّ(الاسم – الفعل –الحرف –الأداة)، معتمدًا على المعنى والصِّيغة ووظيفة اللفظ في الكلام في تقسيمه للكلام؛ لمُواجهة البُعد المنطقي الذي غلب على النَّحوِيِّين في دراساتهم النَّحوِيَّة والصَّرفِيَّة.[1] ثم سار على نهجه الدكتور مهدي المخزومي، إلا أنه أبدل الكنايات بالحرف؛ ليصبح تقسيمه (الاسم – الفعل – الكنايات والإشارات[2] –الأداة).

وينحو الدكتور تمام حسان نحوَ تقسيمٍ جديدٍ لأقسام الكلام؛ ليصل إلى سبعة أقسامٍ رئيسيَّةٍ (الاسم –الصفة –الفعل –الضمير –الخالفة –الظرف – الأداة) معتمدًا على أُسسٍ تنظيريَّةٍ في تقسيمه من حيث المبنى (أي من حيث الصورة الإعرابية أو الرتبة أو الصيغة أو الجدول أو الإلصاق أو التَّضام أو الرَّسم الإملائيِّ)، والمعنى (أي من حيث التَّسمية أو الحدث أو الزمن أو التَّعليق أو المعنى الجملي)، ثم قسّم كل قسم رئيسيٍّ أقسامًا فرعيَّة يوضحها الجدول رقم ٨: [3]

(١) أنيس(إبراهيم): من أسرار اللغة، مكتبة الأنجلو المصرية، ط٦، ١٩٧٨م، ص ٢٨٢-٢٩٤.

(٢) يقصد بها الضمائر المتصلة والمنفصلة وأسماء الإشارة والموصولات وأدوات الشرط والاستفهام.

(٣) حسان(تمام): اللغة العربية معناها ومبناها، دار الثقافة، ١٩٩٤م، ص٩٠.

<table>
<tr><th>الأقسام الرئيسية</th><th colspan="2">الأقسام الفرعية</th></tr>
<tr><td rowspan="10">الاسم</td><td colspan="2">الاسم المعين</td></tr>
<tr><td rowspan="4">اسم الحدث</td><td>المصدر</td></tr>
<tr><td>اسم المصدر</td></tr>
<tr><td>اسم المرة</td></tr>
<tr><td>اسم الهيئة</td></tr>
<tr><td colspan="2">اسم الجنس</td></tr>
<tr><td rowspan="3">الميمات</td><td>اسم الزمان</td></tr>
<tr><td>اسم المكان</td></tr>
<tr><td>اسم الآلة</td></tr>
<tr><td colspan="2">الاسم المبهم</td></tr>
<tr><td rowspan="5">الصفة</td><td colspan="2">صفة الفاعل</td></tr>
<tr><td colspan="2">صفة المفعول</td></tr>
<tr><td colspan="2">صفة المبالغة</td></tr>
<tr><td colspan="2">الصفة المشبهة</td></tr>
<tr><td colspan="2">صفة التفضيل</td></tr>
<tr><td rowspan="3">الفعل</td><td colspan="2">فعل ماض</td></tr>
<tr><td colspan="2">فعل مضارع</td></tr>
<tr><td colspan="2">فعل أمر</td></tr>
</table>

الأقسام الرئيسية	الأقسام الفرعية	
الضمير	حضور	تكلم
		خطاب
		إشارة
	غيبة	شخصية
		موصولة
الخالفة	الإخالة (اسم الفعل)	
	الصوت	
	التعجب	
	المدح والذم	
الظرف	الزمان	
	المكان	
الأداة	أصلية	
	محولة	

جدول: ٨ أقسام الكلام العربي-د.تمام حسان

يلاحظ في هذه التَّقسيمات أنَّ المفردة قد تكون قسمًا كلاميًّا واحدًا مثل الاسم أو الصِّفة أو الفعل أو الضمير،... و قد تكون مكونةً من قسمَيْن كلاميَّيْنِ أو أكثر مثل كلمة «عملهم» (اسم وضمير) و«لهم» (أداة وضمير)، كما أنَّ القسم الكلاميَّ الواحد قد يكون مكونًا من مورفيمٍ واحدٍ مثل: قد - هل - في - لكن،...وقد يكون مكونًا من

أكثر من مورفيم مثل الأسماء والأفعال التي تتركَّب من مورفيمات عديدة. [(١)]

وعلى الرُّغْم من هذه المحاولات التي تعدُّ منطلقًا لإرساء التَّجديد في الدَّرس اللُّغويِّ الحديث، وإعادة النَّظر في المسلَّمات التي بُنيَت عليها النَّظرِيَّة النَّحوِيَّة الوصفية، فإنَّها تقف قاصرةً على الإلمام بوصف الظَّاهرة اللُّغوِيَّة؛ لاعتمادها على التعليل والتأويل دون النَّظر في إطار الظَّاهرة اللُّغوية - أعني بالظَّاهِرة هنا النَّصَّ المكتوب لا المنطوق- والنَّاظر في الواقع اللُّغويِّ سوف يجد أنَّ هناك كلماتٍ ذات سلوكٍ متباينٍ، إدراجها تحت أي قسم كلاميٍّ من التَّقسيمات السَّابقة سيفقدها بعضًا من خصائِصها وصفاتها مثل الكلمات المعرّبة أو العلامات وغيرها.

١-١ ما الحاجة لتقسيم الكلام؟

لعلَّ أوَّل مطالب فَهْم المنظومة اللغوية وتحلِيلها أن تُحدَّد عناصرها الذَّرِيَّة أوَّلًا، إذ لا يُكشَف معنى النَّصِّ إلا إذا فُكِّكَتْ أجزاؤه، وحُدِّدَتْ مكوِّناته، فكان الفونيم هو الوحدة الأساسيَّة لبناء الصَّوت الكلاميِّ، والمورفيم هو الوحدة الأساسيَّة لتكوين المباني الصَّرفِيَّة، بينما التَّاجميم أو السَّنتيم هو الوحدة القاعدِيَّة التي تُبْنى عليها العَلاقات التَّركيبيَّة في الجملة باعتباره بناءً لغويًّا متماسكًا - سواء أكان مورفيمًا أو قسمًا كلاميًّا – مساهمًا في تحديد الوظيفة النَّحوِيَّة والمعنى اللُّغوِيِّ، لينتحي بذلك لِمَنْ يَنْشدُونَ أنْ يتعلَّموا العربيَّة والكشف عن أسرارها وأساليبها، أو لمعالجتها حاسوبيًّا.

وقد ذكر الدُّكتور تمَّام حسان أن هناك أسبابًا لجعل تقسيم الكلِم أمرًا ضروريًّا للتَّعبير عن المعاني النَّحوِيَّة، أوجزها فيما يلي: [(٢)]

١-يتوقَّف جزء من المعنى النَّحوِيِّ على البنية الصَّرفِيَّة؛ إذ يتطلَّب الموقع النَّحوِيُّ قالبًا يشغلُه، فالفاعل والمفعول يتوقَّفان على كون البنية اسمًا، ويتوقَّف المفعول المطلق على بنية المصدر من مادَّة الفعل، بينما يتوقَّف المفعول لأجله على بنية المصدر من غير مادَّة الفعل، وكذلك تتطلَّب كل الفراغات الوظيفيَّة شواغلَ تشغلها، ومن ثمَّ كانت الحاجة إلى تعيين الشواغل بتقسيم الكلِم إلى فئاتٍ تُعَبِّر عن تلك الفراغات الوظيفيَّة.

٢-تتَّضح المعاني الصرفيَّة الهامَّة بتقسيم الكلِم، منها المسمَّى، والموصوف بالحدث، واجتماع الزَّمن والحدث،..إلخ.

---

(١) ينظر: شمس الدين(جلال): الأنماط الشكلية لكلام العرب، الجزء الأول، ص٦٨-٧٠.

(٢) الساقي (مصطفى): أقسام الكلام العربي من حيث الشكل والوظيفة، مكتبة الخانجي بالقاهرة، ١٩٧٧م، ص ١-١٨.

٣-الفصل بين أقسام الكلِم يجعل اللُّغة في مأْمَنٍ من اللَّبس في فهم بعض الأقسام التي تنقل في الاستعمال إلى أقسام أخرى، كنقل الفعل والوصف إلى العلَمِيَّة والاسم إلى الظرفيَّة،..إلخ.

٤-بيان المعاني الوظيفيَّة للمبنى الواحد، ويتَّضح ذلك في كثرة صيغ الاسم والوصف، وتعدُّد صور الضمير، وكذلك الظَّرف والأداة.

٥-من خلال تقسيم الكلم يُمْكِنُنا تحديد المعرَب والمبنى، ومن ثَمَّ الانتفاع بقرينة الإعراب في الكشف عن المعنى النَّحوِيِّ.

٦-تحديد المباني التي تُحدِّد القرائن اللَّفظيَّة المركَّبة في النَّمط الجُمَلِيِّ مثل قرينة الرَّبط التي تتطلَّب مبنيَيْنِ: هما الضَّمير والمرجع، وكذلك الرَّبط بالمطابقة يتطلَّب مبنيَيْنِ متطابقَيْنِ أو أكثر،...إلخ ويتأَتَّى ذلك في ضوء تقسيم الكَلِم.

٢. العنونة بالأقسام الكلامية POS Tagging

هي عمليَّة تعيين الأقسام الكلامِيَّة وما تحمله من سمات صرف-نحوِيَّة لكلِّ كلمة منفردة بمعزلٍ من سياقها الإعرابِيِّ في النَّصِّ، وذلك بإلحاق كل مفردة برمز Tag أو عدَّة رموز تعبِّر عن القسم الكلاميِّ وما يحتويه من مورفيمات أخرى.[١]

وتقتضي جُلُّ أنظمة تطبيقات معالجة اللُّغة الطبيعيَّة واسترجاع المعلومات استخلاصَ السِّمات النَّحوِيَّة البدائيَّة التي تحملها مفردات النُّصوص؛ لتُعينها على تعيين القوالب التَّنظيمِيَّة لِلُّغة[٢]، إذ تعتبر هذه السمات هي حجر الأساس في منحى التَّحلِيل اللُّغوِيِّ الَّذي يتطلَّب تحليل النُّصوص إلى عناصرها الأوَّلِيَّة باعتبارها أجزاءً لِلكَلام؛ وذلك لسهولة التَّصنيف والتَّنظيم والتَّرتيب، والبحث عن القوالب والسِّمات المحدَّدة[٣].

تمثل العَنْوَنة بالأقسام الكلامِيَّة دليلًا أولًا إلى معرفة الملامح اللسانية الأساسية للعديد من التَّطبيقات الحاسوبية الَّتي تمتزج بالبرمجيَّات من ناحية، وبالشَّبكة العنكبوتيَّة من ناحية أخرى؛ ومنها:

(1) Van Halteren, H. (1999). Syntactic Wordclass Tagging. Springer-Science+Business Media.B.V. University of Nijmegen. P.3. & Attia, M. (2004).Theory and Implementation Of a Large-Scale Arabic Phonetic Transcriptor, and Applications. PhD dissertation, Faculty of Engineering, Cairo University, P. 42.

(2) Aliwy, A. (2013). Arabic Morphosyntactic Raw Text Part of Speech Tagging System. PhD dissertation Under the Supervision of Prof. dr hab. Jerzy Tyszkiewicz, Faculty of Mathematics, Informatics and Mechanics, University of Warsaw. P.7.

(3) Van Halteren, H. (1999). Ibid. P.7.

• استرجاع المعلومات Information Retrieval

تزايدت المعلومات بصورة متسارعة في ظل تنامي الثَّورة المعلوماتيَّة العارمة، وانتشار الشَّبكة المعلوماتيَّة التي تتعامل مع العديد من المعارف مثل الكتب والدَّوريَّات والنَّشرات والصُّحف والأبحاث العلميَّة، وغيرها من دوائر المعارف الأخرى؛ مما أدَّى إلى صعوبات جمَّة في استخلاصها وتنقيتها من الخضم المعلوماتيِّ، فكان لزامًا على المعلوماتيين مواجهة هذه التَّحديَّات ببناء نظم للمعلومات قادرة على استخلاصها وتجميعها، تنطلق من المعطيات اللُّغويَّة في تشخيص الكلمات المفتاحية؛ للكشف عن محتوى النَّصِّ.

وتتجلّى تلك المعطيات اللُّغويَّة في التَّعرُّف على خصائص كلمات النَّصِّ من حيث أجزاؤها الكلاميَّة وسماتها النَّحويَّة؛ لتعيِّن الكلمات المِفتاحية Key words لهذا النَّصِّ الَّتي يمكن من خلالها الكشف عن مضمون النَّصِّ، وتحديد مواطن المعلومات الهامَّة آليًّا. [(١)]

• التحليل النحوي Syntactic Parsing

تمهّد العَنْوَنة الكلاميَّة للتَّعرُّف على العَلاقات النَّحوِيَّة، إذ هِيَ من أهمِّ المُدْخَلات الابتدائِيَّة للتَّحلِيل النَّحوِيِّ، حيث لا يُمكنُنا معرفة الوظيفة النَّحويَّة الدَّالة على الكلمة في النَّصِّ (فاعل، مفعول، مضاف إليه...) دون معرفة نوعها الكلاميِّ (اسم، فعل،...) [(٢)]، كما تدفع اللَّبس التَّركيبيِّ في عرض التَّمثيل النَّحوِيِّ للجمل الذي ترسى دعائمه المركبات النحوية Phrases كما سنعرض لاحقًا، علاوة على أنها خُطوة أساسيَّة في الاستدلال على التَّشكيل الآليِّ لِلكلمات من خلال خواصها الصَّرف – نحويَّة. [(٣)]

• التدقيق الإملائي والنحوي Spelling and Grammar checking

تتطلَّب العديد من نُظُم المدقِّقات الإملائيَّة والنَّحويَّة التَّعرُّف على الأنماط الشَّكليَّة للكلمة أو الجملة من حيث الأقسام الكلامية والعَلاقات فيما بينها؛ بغية الانطلاق منها في التَّحقُّق من الأخطاء الإملائيَّة والتَّركيبيَّة. مثلًا على المستوى التَّركيبيِّ يستطيع النِّظام

(1) Alqrainy, sh. (2008). A Morphological – Syntactical Analysis Approach For Arabic Textual Tagging. PhD dissertation, School of Computing Faculty of Computing Sciences and Engineering De Montfort University. P.18.

(2) Attia, M. (2004).Ibid.P.42.

(3) Attia, M. (2004).Ibid.P.48.

أن يُدرِك أنَّ مخالفة النَّمط التَّركيبيِّ الَّذي ينطوي تحت «جر الاسم بعد حروف الجر» يعدُّ من قبيل الخطأ، وذلك من خلال معرفته – السَّابقة – بالنَّمط الشَّكليِّ للكلام، أمَّا على المستوى الإملائي، إذا استطاع النِّظام أن يميِّز بين مواضع الفعل والاسم، قطعًا سيكتشف الخطأ، ثم يقدِّم بديلًا مقترحًا لكلمة «سعى» في عبارة «سعى العمرة»؛ لتصبح «سعي» بالياء لا بالألف المقصورة؛ لكونها اسمًا هنا.

• تخليق الكلام Speech synthesis

تستند معظم أنظمة تخليق الكلام أو تحويل النَّصِّ المكتوب إلى المنطوق (TTS) Text-to-Speech في بنائها على أساليب التَّعلُّم الإحصائيِّ الَّتي يتم بموجبها تكوين الكلمات من وحدات صوتيَّة صغرى سواء أكانت فونيمات أو ديفونات[١] أو مقاطع صوتية أو مورفيمات صرفيَّة حسب النَّهج المتَّبع لِلنِّظام[٢]، والَّذي يكون ناتجًا – غالبًا – عن مدى ثراء Richness إحدى هذه الوحدات في المدوَّنة الصَّوتيَّة المستخدمة.

ويتم هذا التَّوليد أو التَّخليق بتحويل سلسة الحُروف المكتوبة إلى سلسلة من الفونيمات، ثم اختيار الوحدات الصَّوتيَّة (ديفونات، أو ترايفونات، أو مقاطع صوتيَّة، أو مورفيمات صرفيَّة) من المدوَّنة الصَّوتيَّة بناءً على احتمالات الوحدات المتجاورة؛ لتركيبها على تلك الفونيمات الكتابيَّة فيما يعرف بالتَّوليد التَّسلسليِّ concatenative synthesis.[٣]

وتتراوح دقَّة الأداء الصَّوتيِّ للنِّظام حسب ملاءمة الوحدات الصوتيَّة ونوعها للفونيمات الكتابيَّة. ولتحقيق أداء أعلى من جودة الصَّوت، يمكن استخدام المقاطع الصَّوتيَّة أو المورفيمات كعنصر صوتيٍّ لتوليد الكلمات آليًّا، فهي عناصر تحمل في طيَّاتها معطيات أكثر عن وضع الفونيم داخل الوحدات الصوتية. [٤]

ثم تأتي الأقسام الكلاميَّة لتزيد الأمر وضوحًا ودقَّةً، حيث تعدُّ الإشكالية في اختيار المورفيم الصَّوتيِّ من بين العديد من صور المورفيمات الصَّوتيَّة المختلفة الَّتي تشملها العَيِّنة الصَّوتيَّة؛ ليتلاءم مع سلسلة الفونيمات الكتابيَّة، بينما تحديد القسم الكلاميِّ – مسبقًا – للكلمات المكوَّنة من الوحدات الصَّوتيَّة وللكلمات الَّتي سيتم التعرُّف عليها صوتيًّا، سيصوِّب اختيار الوحدات الصَّوتيَّة في مظانِّها؛ لطبيعة ملاءمة النُّطق الصَّوتيِّ

(١) الديفون: هو المنطقة الزمنية الواقعة بين نهاية الفونيم الأول وبداية الفونيم الثاني.

(2) Jurafsky.D, Martin.J. (2007). Ibid. P.316-326.

(3) Jurafsky.D, Martin.J. (2007). Ibid. P.275.

(٤) علي (نبيل):اللغة العربية والحاسوب، تعريب، ١٩٨٨م، ص ٤٤٦.

للوحدات الصَّوتيَّة المشتركة في القسم الكلاميِّ الواحد، ومن ثمَّ تتلاءم الوحدات الصَّوتيَّة مع بعضها البعض في تسلسلٍ مترابطٍ؛ ليخرج الصَّوت بشكلٍ طبيعيٍّ. [١]

• الترجمة الآلية Machine Translation

تعد الترجمة الآليَّة بين اللُّغات الطَّبيعيَّة إحدى الغايات النَّهائيَّة الَّتي تصبُّ فيها معظم روافد نُظُم التَّحليل والتَّركيب اللُّغويَّيْن[٢]، «وقد تحقق العلماء، إثر تجارب مريرة أنهم بعيدون عن بلوغ هذه الغاية»[٣] على الرُّغْم ما حققَّته التَّرجمة الآليَّة الإحصائيَّة Statistical Machine Translation (SMT) من نتائج لا بأس بها، فإن التَّنظير اللِّسانيِّ لم يبلغ حدَّ الكفاية في فهم إشكاليَّة اللُّغة إلى الآن، على أمل في إقحام العلوم الإنسانية المعرفية تلك الإشكالية في ظل بزوغ مجتمع المعرفة؛ للكشف عن الحدس الذي يمتلكه الإنسان وفطرته اللُّغويَّة.

وقد حظيت التَّرجمة الآليَّة الإحصائيَّة مؤخَّرًا باهتمام بالغ من قِبَل العديد من المؤسَّسات والشَّركات ووزارات الدِّفاع العالميَّة، نظرًا لدقة النتائج التي تحققها مقارنةً بالترجمة الآلية المعتمدة على المنهج القاعدي Rule-based .

ويمكن استخدام المدوَّنات اللُّغويَّة المتوازية[٤] Parallel Corpora الموصَّفة بالأقسام الكلامية في بناء النماذج الإحصائية؛ لزيادة دقة نظم الترجمة الآلية. [٥]

• صناعة المعجمات Lexicography

"كان المعجم مستوىً من مستويات النِّظام اللُّغويِّ في الرُّؤية اللِّسانيَّة المتعارفة الذي يُعنى بتصنيف المفردات إلى مداخل لغوية وتعريف محتواها، ثم أصبح بما هو كتاب أو قرص مدمج مجمعَ الأدلَّة على تحليل اللُّغة وتركيبها في اللِّسانيَّات الحاسوبيَّة"[٦] في

---

(1) Alqrainy, sh. (2008). Ibid. P.18.

(٢) على (نبيل):اللغة العربية والحاسوب، ص ١٤٥.

(٣) الموسى (نهاد): العربية نحو توصيف جديد في ضوء اللسانيات الحاسوبية، المؤسسة العربية للدراسات والنشر، ط١، ٢٠٠٠، ص ٥٤.

(٤) المدونات المتوازية هي مجموعة من نصوص لغة من اللغات وترجمات هذه النصوص في لغة أخرى. ينظر:
Lüdeling, A & KytÖ, M. (2008). Ibid. 275.

(5) Alqrainy, sh. (2008). Ibid.P.18-19.

(٦) الموسى (نهاد): العربية نحو توصيف جديد في ضوء اللسانيات الحاسوبية، ص٢٥٢.

ضوء تعدد أغراض الصناعة المعجمية التي تقتضي العديد من المعلومات اللغوية وفق ما تحدده طبيعة المعجم والغاية منه.

وتعددت أشكال الصِّناعات المعجميَّة في ظِلِّ استخدام المدوَّنات اللُّغويَّة الَّتي تعكس الواقع اللُّغويِّ لِلُّغة المدروسة[1]؛ لتشملَ معاجم الموضُوعات والعلوم والفنون والآداب، والمعاجم التَّكراريَّة وشبكات الكلمات، وغيرها من أشكال الصِّناعات المُعْجَمِيَّة الَّتي تمضي في استشفاف معطياتها بتعيين الأقسام الكلاميَّة لمفرداتها.

• تحليل المشاعر Sentiment Analysis

مع تعاظم استخدام شبكات التَّواصُل الاجتماعيِّ وتناميها في ظلِّ بزوع مجتمع المعرفة، تعدَّدت موارد المعلومات التي لعبت دورًا كبيرًا في ارتفاع مؤشرات قياس الأداء الاقتصاديِّ والسِّياسيِّ في إطار تدخُّل تِقنيات اللُّغة لاقتصاد تلك المعرفة المتمثلة في محتويات مِلفَّات مستخدمي شبكات التَّواصُل الاجتماعيِّ مثل (الاسم، الجنس، العمر، المنطقة السَّكنيَّة، الاهتمامات، التَّدوينات، التَّعليقات، الإعجاب بالتَّدوينات). [2]

وتعدَّدت الأغراض من اقتصاد تلك المعلومات بين معرفة معدلات الاستهلاك، وتوقع أنماط الطلب، واحتمالات الكسب والخسارة، واستهداف بعض الإعلانات، واستطلاع الرَّأي، والاستدلال على الشَّخصيَّة، وغيرها من الأهداف الأخرى. [3]

وتنحو تطبيقات تحليل المشَاعِر إلى تصنيف نصوص (تدوينات) مستخدمي شبكات التَّواصُل الاجتماعيِّ ما بين كونها إيجابيَّة positive أو سلبيَّة negative أو محايدَة neutral لقياس مدى بعض التَّوجُّهات السِّياسيَّة أو الاقتصاديَّة أو الاجتماعيَّة[4]...، ويتأتَّى ذلك التَّصنيف في ضوء الأدلَّة اللُّغوية التي تنطلق من الكلمات والعبارات المعبّرة عن تلك العواطف. وتقتضي آليات التصنيف Classfiers استقراء تلك الأدلة من خلال تجلي نواظم نصوص تلكم التدوينات بالسمات الصرف – نحوية. [5]

---

(1) Lüdeling, A & KytÖ, M. (2008). Ibid. P.136.

(2) Sakaki, Sh & Miura, Y & Ma, X & Hattori, K & Ohkuma, T. (2014). Twitter User Gender Inference Using Combined Analysis of Text and Image Processing. In Proceedings of the 25th International Conference on Computational Linguistics. Dublin, Ireland. P.54.

(3) Sakaki, Sh & Miura, Y & Ma, X & Hattori, K & Ohkuma, T. (2014). Ibid.P.54.

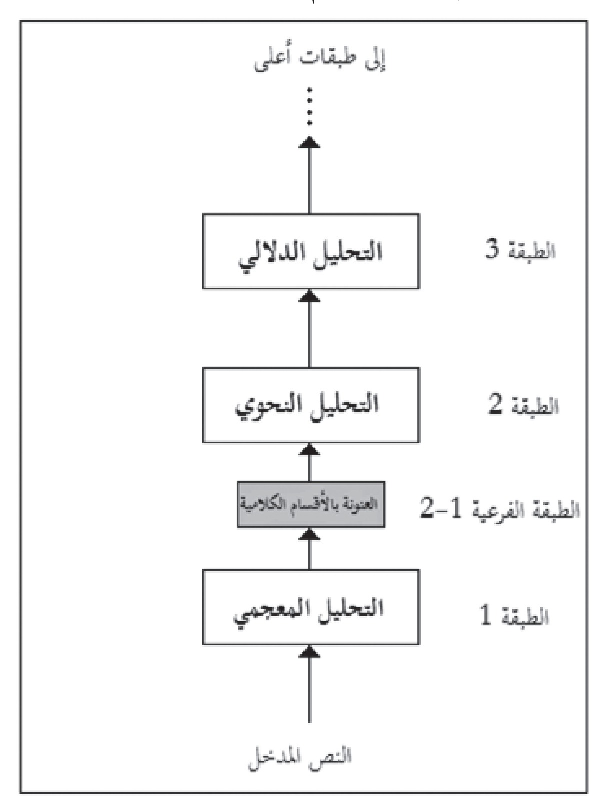
(4) Cambria, E & Poria, S & Bisio, F & Bajpai, R & Chaturvedi, I. (2015). The CLSA Model: A Novel Framework for Concept-Level Sentiment Analysis. In Proceedings CICLing 2015, Egypt. Springer International Publishing Switzerland 2015. A. Gelbukh (Ed.). Part II, P.7.

(5) Peng, B & Li, J & Chen, J & Han, X & Xu, R & Wong, K. (2015). Trending Sentiment-Topic Detection on Twitter. In Proceedings CICLing 2015, Egypt. Springer International Publishing Switzerland 2015. A. Gelbukh (Ed.). Part II, P.71.

• التَّعرُّف على كينونة الاسم Named Entity Recognation

يندرج هذا التَّطبيق NER تحت فرع استخلاص المعلومات Information Extraction حيث يعنى باستخلاص الفئات الاسمية من النصوص مثل أسماء الأعلام، أسماء الهيئات والمنظمات، الأسماء الدالة على الزمان، الأسماء الدالة على المكان، أسماء الأعداد،...إلخ والتي يُنطلق منها – فيما بعد – في بناء أنظمة إحالات الضمير anaphora resolution لإحالة الضمائر إلى أسمائها في النص.[١]

فالعَنْوَنة بالأقسام الكلامِيَّة بمثابة العمود الفقري للعديد من التطبيقات اللسانية التي تتقارب من تناظر الأداء الإنساني بما يتعلق بالملكة اللغوية، ويوضح الشَّكل رقم ١٧ موقعَ العَنْوَنة الكلاميَّة في طبقات سلَّم معالجة اللُّغة الطبيعيَّة:



الشكل:١٧ العنونة الكلامية في طبقات سلم معالجة اللغة الطبيعية[٢]

(1) Jurafsky.D & Martin.J. (2007). Ibid. P.832.

(2) See: Attia, M. (2004).Ibid. P.34.

يلاحظ من خلال الشَّكل السابِق أنَّ العَنْوَنة بالأقسام الكلامِيَّة تتوسَّط بين التَّحلِيل المعجميِّ Lexical Analysis والتَّحلِيل النَّحوِي Syntactic Analysis في طبقات معالجة اللُّغات الطبيعِيَّة؛ لتكونَ مَدْخلًا للتَّحلِيل النَّحوِيِّ الذي يُناط به توفير المعطيات اللَّازِمة للتَّحلِيل اللُّغويِّ الأعمق لِلْفهم الأتوماتي للنُّصوص اللُّغوِيَّة، فضلًا عن كونها منطلقًا أساسيًّا لتطبيقات التحليل اللساني.

ولما كان المأمول من عَنْوَنة المدوَّنات اللُّغوِيَّة بالأقسام الكلامِيَّة أنْ تكون موردًا لمعالجة اللُّغة الطبيعيَّة، كان لابدَّ أن تُحدَّد قبل الشُّروع في العَنْوَنة منهجيَّةُ الأقسام الكلامِيَّة الَّتي تُصنِّف مفردات مادَّة المُدوَّنة اللُّغوِيَّة، وذلك باستقراء الخصائص الصَّرف-نحوِيَّة التي تتناسب والهدفَ المنشود، ثم تجريدها على هيئة عدد محدود من الرُّموز- تسمَّى Tag set- التي تعبِّر عن تلكم السِّمات اللُّغوِيَّة، وتتفاوت عدد السمات حسبما تهدف إليه.[1]

وبدأت تتبلور تلك الأقسام الكلاميَّة لِلُّغة الإنجليزِيَّة في مطلع النِّصف الثاني من القرن العشرين على هيئة فئات من السِّمات الصَّرف- نحوِيَّة تنظر لِلنَّصِّ باعتباره الظاهرة موضع الدراسة، فتصدرت مُدَوَّنة بروان لِلُّغة الإنجليزِيَّة تلك الفئات؛ ليصل عدد سماتها ٨٧ سمة، بينما تقلُّ عدد سمات بنك بنسلفانيا النَّحوِيِّ؛ لتصل إلى ٤٥ سمة، وتتوسَّطهما عدد السِّمات المستخدمة في مشروع CLAWS والتي بلغت ٦١ سمة.[2]

إن غنى اللُّغة العربيَّة صرفيًّا أدّى إلى وجود العديد من المورفيمات التي يصعب العمل عليها جميعًا، فضلًا عن أن كثيرًا من الباحثين في معالجة اللُّغة الطبيعِيَّة يفضِّلون العمل على فئات السِّمات اللُّغوِيَّة الأصغر حجمًا؛ حتى يمكن التَّنبُّؤ بها بدِقَّة في ضوء استخدام أساليب التَّعلُّم الإحصائِيِّ.[3]

وقد تعدَّدت منهجيات الأقسام الكلاميَّة للُّغة العربيَّة التي تُصنّف مفردات النص في ضوء وصف الواقع اللغوي، منها:

(1) Sawalha, M. (2011). Open-source Resources and Standards for Arabic Word Structure Analysis: Fine Grained Morphological Analysis of Arabic Text Corpora. Submitted in accordance with the requirements for the degree of Doctor of Philosophy From The University of Leeds School of Computing.. P.96.

(2) Sawalha, M. (2011). Ibid. P.96.

(3) Habash , N. (2010). Ibid. P.79.

٢-١ فئة خوجة الكلامية (Khoja Tag Set)

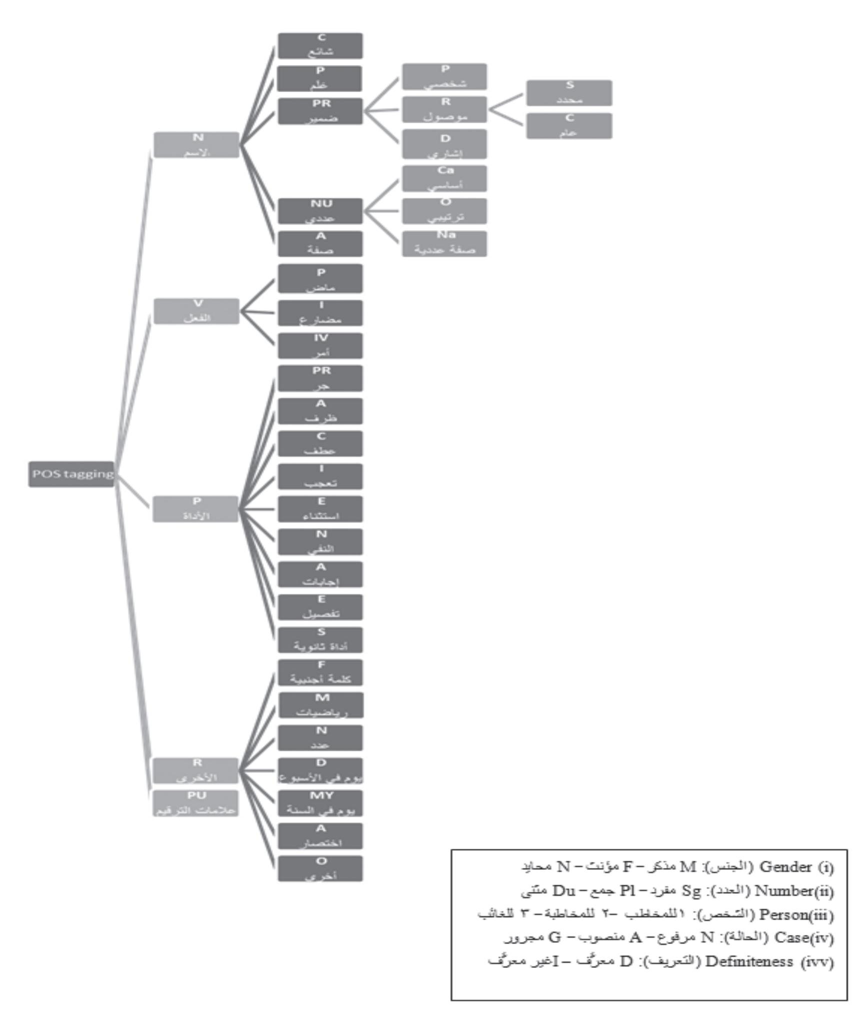
تعتبر الباحثة شيرين خوجة[١] صاحبة أول محاولة جادَّة في تصميم فِئَةِ الأنواع الكلاميَّة POS Tagset لِلنَّصِّ العربيِّ، حيث قامت بوضع منهجية مغايرة للأجزاء الكلاميَّة غير الَّتي تسير عليها اللُّغات الأوروبيَّة لأسباب قد ذكرتْها في ورقتها المعَنْوَنة بـ«مُعَنْوِن الأجزاء الكلاميَّة لِلنَّصِّ العربيِّ» منها: أنَّ اللُّغة العربيَّة لها طبيعة خاصَّة، إذ تنتمي لفصيلة اللُّغات السَّاميَّة، بَيْنَما المعايير الأوربيَّة لعنونة المُدَوَّنات اللُّغوِيَّة AGLES[٢]صُمِّمَت خصيصًا للُّغات الهندوأوروبيَّة، كما أنَّ هذه المعايير لا تُقَدِّم بعض المعلومات الصَّرف-نحوِيَّة التي تتفق وحالَ العربيَّة مثل البناء والإعراب، فانطلقتْ من النَّحو العربيِّ التَّقليديِّ Traditional grammar - كما يطلقون عليه الوصفيُّون- واستقراء الواقع اللُّغويِّ المعاصر؛ لتصل في تقسيمها إلى خمسة أقسام رئيسيَّةٍ (الاسم – الفعل – الأداة – أخرى – علامة الترقيم)، وكلُّ قسم من تلكم الأقسام يحتوي على تقسيمات فرعية[٣] كما بالشكل رقم ١٨.

---

(١) تعدّ الباحثة الدكتورة شيرين خوجة من رواد معالجة اللغة العربية آليا، وقد نالت الدكتوراه في معالجة اللغة الطبيعية من جامعة لانكاستر عام ٢٠٠٣م، ، وتعمل حاليا في وظيفة أستاذ مشارك في قسم علوم الحاسب جامعة باسيفك Pacific بالولايات المتحدة الأمريكية، ومن اهتماماتها العملية في مجال اللغويات الحاسوبية: التجذيع، وقد طورت معنونا آليا للأقسام الكلامية للنص العربي APT، وكذلك صناعة المدونات اللغوية للغة العربية. وقد صممت فئة الأجزاء الكلامية بمساعدة اللغوي جيري نيولز Gerry Knowles أستاذ اللغويات بجامعة لانكاستر. ينظر:
Shereen Khoga. (2015). http://zeus.cs.pacificu.edu/shereen/index.html

(٢) تقدم المجموعة الاستشارية الخبيرة في معايير هندسة اللغة The Expert Advisory Group on Language Engineering Standards (EAGLES) المبادئ العامة التوجيهية لترميز نصوص المدونات اللغوية. ينظر:
EAGLES. (2015). http://www.ilc.cnr.it/EAGLES/browse.html

(3) Khoja, Sh. (2001). APT: Arabic Part-of-speech Tagger. In Proceedings of the Student Workshop at the Second Meeting of the North American Chapter of the Association for Computational Linguistics (NAACL2001), Carnegie Mellon University, Pittsburgh, Pennsylvania.



شكل ١٨ : فئة خوجة الكلامية

وبهذا يصل عدد السِّمات المنفردة التي أقرَّتها خوجة إلى ٥١ سمة، متضمّنة علامات الفصائل النَّحوِيَّة (التَّعريف – العدد – الجنس – الحالة – الشَّخص)، وعندما تتكامل فيما بينها تصل إلى ١٧٧ سمة، تتوزع على الأقسام الرئيسيَّة كالآتي: ١٠٣ اسما، ٥٧ فعلا، ٩ أدوات، ٧ أخرى، ١ علامة الترقيم.[١] ويوضح الجدول رقم ٩ مثالًا لعَنْوَنة خوجة بالأقسام الكلاميَّة لجملة «تنفيذًا لتوجيهات خادم الحرمين الشريفين».

(1) Khoja, Sh. (2001). An Arabic Tagset for the Morphosyntactic Tagging of Arabic. In Proceedings of Corpus Linguistics, Lancaster University, Lancaster, UK, March. P.3.

الكلمة	مورفيمات الأجزاء الكلامية	المصطلح العربي
تنفيذًا	NCSgMI	اسم_عام_مفرد_مذكر_غير معرف
لتوجيهات	PPr'NCSgMI	أداة_جر_ اسم_عام_مفرد_مذكر_غير معرف
خادم	NCSgMI	اسم_عام_مفرد_مذكر_غير معرف
الحرمين	NCDuMD	اسم_عام_مثنى_مذكر_معرف
الشريفين	NCDuMD	اسم_عام_مثنى_مذكر_معرف

الجدول:٩ عنونة خوجة بالأقسام الكلامية - نموذج من مدونة خوجة[١]

يلاحظ أن عدد السِّمات يختلف في النَّصِّ المجزَّأ للزَّوائد Clitics - بنوعيها - عن غيره، إذ نجد في المثال السابق - غير مجزَّأ من ناحية الزَّوائد - تقلُّصَ عدد السِّمات إلى خمس سمات تندرج جميعهم تحت القسم الكلاميِّ الاسم N، أمَّا في حالة التَّجزِئة بالزَّوائد فيزيد عدد السِّمات؛ لأنَّ بعض الزَّوائد مثل الأدوات تعتبر قسمًا كلاميًّا.كما يلاحظ أن هذه المنهجية تفصل بين الأجزاء الكلامية في الكلمة بعلامة (').

ومما يؤخد على هذه الفئة أن الهدف منها لم يكن محددًا، وإن كانت غايتها أن تكون شاملةً لتطبيقات معالجة اللغة الطبيعية، فهي لم تستوعب كل الخصائص الصرف-نحوية للغة العربية التي تعين على تمييز مفردات النص، مثل: عدم تمييز وضعية السوابق واللواحق في الكلمة، عدم تعيين رموز تدل على قطعية احتمال حالة النصب أو الجر للضمائر الملتصقة بالكلمات.

٢-٢ فئة باكولتر الكلامية (Buckwalter Tag Set)

قام الأمريكيُّ تيم باكوالتر[٢] Tim Buckwalter بتطوير فِئةٍ جديدةٍ للأنواع الكلاميَّة لِلنَّصِّ

(1) Khoja, Sh. (2001). Ibid.P.7.

(٢) أستاذ اللغويات الحاسوبية بجامعة ماريلاند بالولايات المتحدة الأمريكية، له العديد من الإسهامات في مجال معالجة اللغات الطبيعية، منها على سبيل المثال لا الحصر: التحليل الصرفي للغة العربية، صناعة المعجم العربي، النقل الكتابي.

العربيِّ؛ بهدف إدراجها في محلِّله الصَّرفيِّ BAMA[(١)]، تنطلق في ظاهرها من ثمانية أقسام رئيسيَّةٍ، وهي(الاسم – الفعل – الصفة – الظرف – الحرف – الضمير – الأداة –الأخرى) كما ورد في تطبيقه لتلك الفئة الكلاميَّة، ويوضِّح الشَّكل رقم ١٩ مثالًا من مخرجات Outputs المحلِّل الصَّرفيِّ BAMA –الذي اعتمد على فئة الأقسام الكلاميَّة في تحديد سمات أجزاء الكلمة– لجملة «تم إعداد الوثائق المتوفرة بكثرة حول أول رحلة طيران عثمانية فوق البلاد العربية».

```
INPUT STRING: تم
LOOK-UP WORD: tm
     Comment:
* SOLUTION 1: (tam~) tam~/VERB_PERFECT
     (GLOSS):  + conclude/take place +
INPUT STRING: اعداد
LOOK-UP WORD: AEdAd
     Comment:
  SOLUTION 1: (>aEodAd) >aEodAd/NOUN
     (GLOSS):  + numbers/issues +
* SOLUTION 2: (<iEodAd) <iEodAd/NOUN
     (GLOSS):  + preparation +
INPUT STRING: الوثائق
LOOK-UP WORD: AlwvA}q
     Comment:
* SOLUTION 1: (AlwavA}iq) Al/DET+wavA}iq/NOUN
     (GLOSS): the + documents/charters +
INPUT STRING: المتوفرة
LOOK-UP WORD: Almtwfrp
     Comment:
* SOLUTION 1: (Almutawaf~irap) Al/DET+mutawaf~ir/ADJ+ap/NSUFF_FEM_SG
     (GLOSS): the + available/abundant + [fem.sg.]
INPUT STRING: ب
LOOK-UP WORD: b
     Comment: Separated
* SOLUTION 1: (bi-) bi-/PREP
     (GLOSS): by/with
INPUT STRING: كثرة
LOOK-UP WORD: kvrp
     Comment:
* SOLUTION 1: (-kavorap) -kavor/NOUN+ap/NSUFF_FEM_SG
     (GLOSS): abundance/frequency + [fem.sg.]
INPUT STRING: حول
LOOK-UP WORD: Hwl
     Comment:
* SOLUTION 1: (Hawola) Hawola/PREP
     (GLOSS):  + about/around +
  SOLUTION 2: (Haw~al) Haw~al/VERB_PERFECT
     (GLOSS):  + change/convert/switch +
  SOLUTION 3: (Hawol) Hawol/NOUN
     (GLOSS):  + power +
INPUT STRING: أول
LOOK-UP WORD: >wl
     Comment:
  SOLUTION 1: (>aw~al) >aw~al/VERB_PERFECT
     (GLOSS):  + explain/interpret +
* SOLUTION 2: (>aw~al) >aw~al/ADJ
     (GLOSS):  + first +
  SOLUTION 3: (>uwal) >uwal/ADJ
     (GLOSS):  + first +
INPUT STRING: رحلة
LOOK-UP WORD: rHlp
     Comment:
* SOLUTION 1: (riHolap) riHol/NOUN+ap/NSUFF_FEM_SG
     (GLOSS):  + journey/career + [fem.sg.]
INPUT STRING: طيران
LOOK-UP WORD: TyrAn
     Comment:
* SOLUTION 1: (TayarAn) TayarAn/NOUN
     (GLOSS):  + airline/aviation +
INPUT STRING: عثمانية
LOOK-UP WORD: EvmAnyp
     Comment:
  SOLUTION 1: (EuvomAniy~ap) EuvomAniy~/NOUN+ap/NSUFF_FEM_SG
     (GLOSS):  + Ottoman + [fem.sg.]
* SOLUTION 2: (EuvomAniy~ap) EuvomAniy~/ADJ+ap/NSUFF_FEM_SG
     (GLOSS):  + Ottoman + [fem.sg.]
INPUT STRING: فوق
LOOK-UP WORD: fwq
     Comment:
* SOLUTION 1: (fawoq) fawoq/PREP
     (GLOSS):  + above/over +
  SOLUTION 2: (fawoq) fawoq/NOUN
     (GLOSS):  + top/upper part +
INPUT STRING: البلاد
LOOK-UP WORD: AlblAd
     Comment:
* SOLUTION 1: (AlbilAd) Al/DET+bilAd/NOUN
     (GLOSS): the + (native) country/countries +
INPUT STRING: العربية
LOOK-UP WORD: AlErbyp
     Comment:
```

الشكل:١٩ مخرجات محلل باكولتر الصرفي[(٢)]

(١) محلل باكولتر الصرفي للغة العربية Buckwalter Arabic Morphological Analyzer، للاستزادة حول التفاصيل، ينظر:

Buckwalter, T. (2004). Buckwalter Arabic Morphological Analyzer Version 1.0. Linguistic Data Consortium, University of Pennsylvania, 2004. LDC Catalog NO: LDC2004L02.

(2) Sawalha, M. (2011). Ibid. P.101.

يلاحظ أن المحلِّل الصَّرفيّ يطرح جميع التَّحليلات المُمْكِنة للكلمة الواحدة، ثُمَّ يرجّح محررو البنك الشجري العربي تحليلًا واحدًا في نهاية الأمر - حسب ما يؤول إليه السياق- بوضع علامة * أمام الاختيار المحتمل صحته، كما يقسّم الكلمة إلى أجزائها من حيث السَّوابق والجذع واللَّواحق، ثم يضع أمام كلِّ جزءٍ السمةَ التي تنتمي إليه من الفئة الكلاميَّة.

وقد اشتملَت هذه الفئة الكلاميَّة على ٧٠ سمة أساسيَّة منفردة، وعندما تتحد هذه السِّمات الفرعيَّة تصل إلى ١٧٠ سمة، مثال ذلك اتِّحاد السِّمات الثلاث (NSUFF_ FEM_SG) لِلدِّلالة على لاحقة المفرد المؤنَّث، وكذلك(CASE_DEF_ACC) لِلدِّلالة على حالة النَّصب[1]. ويعرض الجدول رقم ١٠ هذه السمات:

المصطلح العربي	المصطلح الانجليزي	الاختصار
الفعل		
فعل	verb	VERB
شبه الفعل	Pseudo-verb	PSEUDO_VERB
الفعل التام	Perfective verb	PV
الفعل المبني للمجهول	Perfective passive verb	PV_PASS
المفعول المباشر	direct object of Perfective verb	PVSUFF_<DO:<PGN
الفاعل المتصل بالفعل	Subject of Perfective verb	PVSUFF_<SUBJ:<PGN

المصطلح العربي	المصطلح الانجليزي	الاختصار
الاسم		
اسم	Noun	NOUN
عدد	Nominal/ cardinal number	NOUN_NUM
كمية او مقدار	Quantifier noun	NOUN_QUANT
الأسماء المشتقة من الأفعال	Deverbal noun	NOUN_VN
اسم علم	Proper noun	NOUN_PROP
الصفة		
صفة	Adjective	ADJ

(1) Habash , N. (2010). Ibid. P.80.

الاختصار	المصطلح الانجليزي	المصطلح العربي
IV	Imperfective verb	الفعل المضارع
IV_PASS	Imperfective passive verb	الفعل المضارع المبني للمجهول
IVSUFF_ <DO:<PGN	Imperfective verb direct object	المفعول المباشر المتصل بالفعل المضارع
<IV<PGN	Imperfective verb prefix	علامات الفعل المضارع
IVSUFF_ <SUBJ:<PGN _<MOOD:<mood	Imperative verb Subject and mood suffix	الفاعل المتصل بالفعل الأمر وحالته الإعرابية
CV	Imperfective (command) verb	الفعل الأمر
CVSUFF_ <DO:<PGN	Imperative verb object	المفعول المتصل بالأمر

الاختصار	المصطلح الانجليزي	المصطلح العربي
ADJ_COMP	Comparative adjective	صفة مقارنة
ADJ_NUM	Adjectival/ ordinal number	العدد الترتيبي
ADJ_VN	Deverbal adjective	الصفة المشبهة
ADVJ_ PROP	Proper adjective	صفة شائعة
الظرف		
ADV	Adverb	ظرف
REL_ADV	Relative adverb	ظرف بصيغة الموصول
INTERROG_ ADV	Interrogative adverb	ظرف بصيغة الاستفهام
الضمير		
PRON	Pronoun	الضمير

المصطلح العربي	المصطلح الانجليزي	الاختصار
الفاعل المتصل بالأمر	Imperative verb subject	CVSUFF_ <SUBJ:<PGN
الحرف		
حرف الجر	preposition	PREP
حرف العطف	Conjunction	CON
الحروف الثانوية المشبهة بالفعل	Subordinating conjunction	SUBJ_CONJ
الأداة		
الأداة	particle	PART
أداة ربط	Connective particle	CONNEC_ PART
أداة التأكيد	Emphatic particle	EMPHATIC_ PART
أداة تنبيه	Focus particle	FOCUS_ PATR
أداة التسويف	Future particle	FUT_PART
أداة استفهام	Interrogative particle	INTERROG_ PART
أداة جزم	Jussive particle	JUS_PART
الضمير الشخصي	Personal pronoun	<PRON_<PGN
ضمير ملكية	Possessive Personal pronoun	PRON_<PGN> _POSS
اسم إشارة	Demonstrative pronoun	DEM_ <PRON_<GN
اسم موصول	Relative pronoun	REL_PRON
اسم استفهام	Interrogative pronoun	INERROG_ PRON
لاحقة اسمية	Nominal suffix	NSUFF<Gen> <Num> <Cas> <<Stt
لاحقة اسمية	Nominal suffix	CASE <Def> <<Cas
محدّد	Determiner	DET
أخرى		
علامة ترقيم	Punctuation	PUNC
اختصار	Abbreviation	ABBREV

المصطلح العربي	المصطلح الانجليزي	الاختصار
أداة نفي	Negative particle	NEG_PART
أداة جواب الشرط	Response conditional particle	RC_PART
أداة حصر	Restrictive particle	RESTRIC_PART
أداة الفعل	Verb particle	VERB_PART
أداة النداء	Vocative particle	VOC_PART
التعجب	Interjection	INTERJ
العلامات اللاتينية	Latin script	LATIN
الكلمة الأجنبية	Foreign word	FOREIGN
الخطأ الكتابي	Typographical error	TYPO
الكلمة المجزأة	Partial word	PARTIAL
الكلمة العامية	Dialectal word	DIALECT

الجدول: ١٠ فئة باكولتر الكلاميّة[١]

ويتمُّ نظم تلك السِّمات المنفردة في شكل متَّجهٍ[٢] من المورفيمات العربيَّة المتابعة للمفردة، مثال ذلك المتَّجه الكلاميُّ (DET+ADJ+NSUFF_FEM_SG+CASE_DEF_ACC) لمفردة "الجميلَةَ" . وقد استخدمتْ هذه الفئة أيضا في عَنْوَنة بنك بنسلفانيا لِلُّغة العربيَّة PATB بالأقسام الكلامِيَّة[٣]. ويوضِّح الجدول رقم ١١ مِثالًا لِعَنْوَنة بنك بنسلفانيا العربيِّ بفئة باكولتر للأقسام الكلامِيَّة لجملة «سعوديون يشكون من عنف أمريكي خلال التحريات عن هجمات ١١ أيلول».

---

(1) Habash , N. (2010). Ibid. P.81.

(٢) المتجة (Vector) هو مصطلح رياضي يُدْرَس – تحديدًا – في الجبر الخطي Linear algebra ويمكن أن يعرّف بأنه خط يتجه من نقطة لأخرى، ويتم تحديد خصائصه بمقدار طول هذا الخط R، ويستخدم مع نظيره المصفوفة Matrix في بناء نظم البيانات وتنسيقها بما يتوافق مع معطيات آليات التعلم الآلي. ينظر:

Anton, H. & Rorres, Ch. (2010). Elementary Linear Algebra. John Wiley & Sons, Inc. 10th edition P.228.

(3) Maamouri, M & Bies, A & Bukwalter, T & Mekki, W. (2004).The Penn Arabic Treebank: Building a Large-Scale Annotated Arabic Corpus. In Proceedings of NEMLAR Conference on Arabic Language Resources and Tools. P.5.

الكلمة	الأقسام الكلامية	المصطلح العربي
سعوديون	NOUN+NSUFF_MASC_PL_ NOM	اسم+مذكر_جمع_مرفوع
يشكون	IV3MP+IV+IVSUFF_SUBJ:MP_ MOOD:I	حرف مضارعة للمذكر المفرد الغائب+فعل مضارع+فاعل متصل_حالته الإعرابية:الرفع
من	PREP	حرف جر
عنف	NOUN+CASE_INDEF_GEN	اسم+غير مضاف _ مجرور
أميركي	ADJ+CASE_INDEF_GEN	صفة+غير مضافة_مجرورة
خلال	PREP	حرف جر
التحريات	DET+NOUN+NSUFF_FEM_ PL+CASE_DEF_GEN	أداة تعريف+اسم+مؤنث_ جمع+مضاف_مجرور
عن	PREP	حرف جر
هجمات	NOUN+NSUFF_FEM_ PL+CASE_DEF_ACC	اسم+مؤنث_جمع + مضاف مجرور
١١	NUM	عدد
أيلول	NOUN_PROP+CASE_INDEF_ GEN	اسم علم+غير مضاف_مجرور

الجدول: ١١ عنونة بنك بنسلفانيا العربي بالأقسام الكلامية [١]

وتَجْدُر الإشارة هنا إلى أنَّ السِّمات الفرعيَّة (المتَّصلة) يختلف عددها عند العَنْوَنة باختلاف نوع النَّصِّ من حيث كونه مجزَّءًا Toknized أو غير مجزَّأ unTokenized، إذ نجد – على سبيل المثال– في فئة السِّمات المستخدمة في بنك بنسلفانيا العربيِّ سِمتي العطف (CON) والضمير (PRON) مستقلَّتين بذاتهما، فلا يتَّحدان مع سمات أخري،

(١) هذا المثال مقتبس من بنك بنسلفانيا الشجري ATB.

وهذا يعود إلى أن بنك بنسلفانيا العربيِّ يتبع نهج فصل الزَّوائد Clitics في النَّصِّ. وقد يصل عدد السِّمات الفرعيَّة (المتَّصلة) أي السِّمات الممكن ردها لأنواع المفردات المختلفة كما في المثال الذي ذكرته آنفا إلى ٥٠٠ سمة، بَيْنَما قد يصل العدد في النَّصِّ غير المجزَّأ إلى ٢٢ ألف سمة.

٢-٣ فئة بييز الكلامية BIES Tag Set

وانطلاقًا من مبادئ التَّعلُّم الآليِّ التي تقتضي إحكام المعطيات اللُّغويَّة، وتستوجب بناءً نمطيًّا مطردًا، قامت آن بييز Ann Bies[(١)] و دان بايكل Dan Bikel باستخلاص فئة مقلَّصة للأقسام الكلاميَّة للنَّصِّ العربيِّ Reduced Tagset (RTS) من فئة باكولتر الكبرى، تماهي فئة الأقسام الكلاميَّة المستخدمة في بنك بنسلفانيا لِلُّغة الإنجليزيَّة، وذلك بهدف تحسين أداء التَّحلِيل النَّحوِيِّ الآليِّ.[(٢)]

وانطلقت بييز في تقسيمه للأقسام الكلاميَّة الرئيسيَّة على غرار تقسيم باكولتر الثُّماني، غير أنها انتهجت نهج فئة الأقسام الكلاميَّة الإنجليزيَّة التي أقرَّتها مُؤَسَّسة شبكة البيانات اللُّغوِيَّة LDC لعَنْوَنة بنك بنسلفانيا الإنجليزيِّ؛ ليصل بعدد سمات تلك الفئة التي أنشأها إلى ٢٥ سمة[(٣)]، وبهذا تَدْخُلُ ضمن الفئات الكلاميَّة الأصغر حجمًا من حيث عدد السِّمات الصَّرف - نحوِيَّة.

على الرُّغْم من تقارُب إحكام تلك الفئة الكلاميَّة، فإنَّ محاكاتها للأقسام الكلاميَّة الإنجليزيَّة أفقدها بعضًا من المعلومات اللُّغوِيَّة التي تعدُّ ملحًا أساسيًّا في استنباط العَلاقات النَّحوِيَّة، مثال ذلك استخدام رمز JJ مع جميع الصِّفات بغض النَّظر عن خصائص هذه الصِّفات سواءً أكانت اسم فاعل أو صيغة مبالغة أو صفة مشبهة أو اسمًا منسوبًا[(٤)]،...، وهذا يرجع إلى أن الفئة الكلاميَّة لِلُّغة الإنجليزيَّة لا توجد بها تلك الخصائِص.

---

(١) باحث مساعد بدرجة « خبير» بمؤسسة شبكة البيانات اللغوية LDC، وتعمل على العديد من المشروعات في العنونة النحوية والدلالية للغة العربية والإنجليزية.

(2) Habash , N. (2010). Ibid. P.80.

(3) Diab, M.T. (2007). Improved Arabic Base Phrase Chunking with a new enriched POS tag set. In Proceedings of the 5th Workshop on Important Unresolved Matters, Association for Computational Linguistics (ACL), Prague. P.89.

(4) Habash , N. (2010). Ibid. P.80.

ويوضح الجدول رقم ١٢ سمات فئة بيز الكلامية المقلصة التي أقرتها مؤسسة شبكة البيانات اللغوية بجامعة بنسلفانيا في عام ٢٠٠٤م.

المصطلح العربي	المصطلح الإنجليزي	الاختصار
الاسم Noun		
اسم شائع (مفرد)	singular common noun or abbreviation	NN
اسم شائع (جمع أو مثنى)	plural/dual common noun	NNS
اسم علم (مفرد)	singular proper noun	NNP
اسم علم(جمع أو مثنى)	plural/dual proper noun	NNPS
عدد	cardinal number	CD
الضمير Pronouns		
ضمير شخصي	personal pronoun	PRP
ضمير ملكية	possessive personal pronoun	$PRP
ضمير وصل	relative pronoun	WP
الصفة		
الصفة	Adjective	JJ
الظرف		
الظرف	Adverb	RB
ظرف بصيغة الاستفهام	relative adverb	WRB

المصطلح العربي	المصطلح الإنجليزي	الاختصار
الحرف		
حرف عطف	coordinating conjunction	CC
حرف جر/ حرف عطف	preposition or subordinating conjunction	IN
الأدوات		
محدد/ اسم إشارة	determiner/demonstrative pronoun	DT
أداة	Particle	RB
حرف جر/ حرف عطف	preposition or subordinating conjunction	IN
الأفعال		
فعل مضارع	active imperfect verb	VBP
فعل مبني للمجهول	passive imperfect/perfect verb	VBN
فعل ماض	active perfect verb	VBD
فعل أمر	imperative verb	VB
الوظيفية		
أداة تعجب	Interjection	UH
الترقيم	Punctuation	PUNC
الفاصلة العددية	NUMERIC_COMMA	NUMERIC_ COMMA
كلمة غير محللة	NO_FUNC	NO_FUNC
كلمة أجنبية	Foreign word	FW

الجدول:١٢ فئة بييز الكلامية المقلصة [١]

(1) Habash , N. (2010). Ibid. P.80.

وبمُلاحظة فئة مورفيمات بييز الكلاميَّة يتبيَّن لنا أنَّها تعتمد بصورة سافِرة على المنظومة الصَّرف – نحويَّة لِلُّغة الإنجليزيَّة دون أن تُراعي خصوصيَّات اللُّغة العربيَّة، مثال ذلك استخدامها لرمز الظَّرف بصيغة الاستفهام WRB الذي يختصّ بتوصيف ظَرْفي اللُّغة الإنجليزيَّة where when ، اللَّذَيْنِ يمتلكان طابع النَّظام الكتابيِّ لصيغة الاستفهام في اللُّغة ذاتها، كما يلاحظ أنها لم تعرْ اهتمامًا لمورفيمات الفصائل النَّحوِيَّة (الجنس–الشخص– التعريف والتنكير) التي تعدُّ - بوصفها أداة لإظهار التَّشابه والتَّرابُط بين المقولات النَّحوِيَّة - مكونًا من مكونات الأقسام الكلامية، وكذلك العلامات الإعرابيَّة.

٢-٤ فئة بادت الكلامية PADT Tag set

أعلن معهد اللُّغوِيَّات التَّطبيقيَّة والصُّورِيَّة بجامعة تشارلز ببراغ Prague في مطلع عام ٢٠٠٤م عن مشروع بناء «بنك شجريّ اعتمادي للُّغة العربيَّة»Prague Arabic Dependency Treebank (PADT) يعتمد في بنيته على البنية الاعتمادية التَّركيبيَّة، بهدف توفير مورد لُغويٍّ للأغراض البحثيَّة والتِّجارِيَّة في معالجة اللُّغة الطبيعيَّة.[١] وقد تضمّنت الخُطَّة المنهجيَّة للبناء هيكلةً تنظيميَّة للمنظومات اللُّغوِيَّة المختلفة قد مثَّلتها على هيئة طبقاتٍ لُغوِيَّةٍ مُتدرِّجةٍ في سلَّم معالجة اللُّغات الطبيعيَّة بدايةً من التَّحليل المعجميِّ مُرورًا بالأجزاء الكلاميَّة وانتهاءً بالتَّحليل النَّحوِيِّ، وما يهمنا في هذا المقام الأجزاء الكلاميَّة.

وقام فريق العمل في مشروع بنك براغ الشجريِّ للُّغة العربيَّة بتطوير فئةٍ كلاميَّةٍ؛ لتكونَ مُنطلقًا مسوَّغًا تجري عليه العَلاقات التَّركيبيَّة مجرىً متسلسلًا في التَّحليل الشَّجريِّ فيما بعد. «وقد تضمّنتْ في محلِّل ElixirFM الصَّرفيِّ الذي استلهم منهجيته من الصَّرف الوظيفيِّ»[٢]؛ بهدف تحليل المدوَّنة اللُّغوِيَّة التي ينطلق منها المشروع اللُّغويُّ تحليلًا شبه آليٍّ أي يحتاج إلى المراجعة والتعديل.

---

(1) Smrž, O & Bielický, V. et al., (2008). Prague Arabic dependency treebank: A word on the million words. In Proceedings of the Workshop on Arabic and Local Languages (LREC) 2008.Marrakech, Morocco. European Language 2008.Marrakech, Morocco. European Language Resources Association.

(2) Habash , N. (2010). Ibid. P.84.

كان الأساس من وراء تصميم هذه الفئة الكلاميَّة هو إمكانية إخضاع السِّمة الكلاميَّة لجزءَيْنِ أساسيَّيْن: هما القسم الكلاميُّ والفصيلة النَّحوِيَّة التي يحملها؛ لتكونَ أكثر تعميمًا وعمقًا من نظيرتها المستخدمة في بنك بنسلفانيا الشجريِّ، إذْ من المفترض أن يُسْتَدَل على السِّمة الكلاميَّة من الصيغة الوظيفيَّة بعكس ما تسير عليه الفئات الأخرى بالاستدلال من الصِّيغة الشَّكلِيَّة لِلْكلمة[1]، ويوضِّح الشَّكلان ٢٠ و٢١ طُرُقَ استمداد السِّمة الكلامِيَّة من الصِّيغة الصَّرفيَّة.

Form	`al-lA-silkIyaTu`	*al-lā-silkīyatu* اَللَّاسِلْكِيَّةُ
Morph	`al >| lA >| FiCL |< Iy |< aT |<< "u"`	
Tag	`A-----FS1D`	
Form	`lA-silkIy`	*lā-silkīy* لَاسِلْكِيّ
Morph	`lA >| FiCL |< Iy`	
Root	`"s l k"`	
Reflex	wireless, radio	
Class	adjective	

الشكل: ٢٠ مخرجات محلل ElixirFM–نموذج "اللَّاسِلْكِيَّةُ" [2]

تستند السِّمة الكلاميَّة في هذا النَّموذج على الصِّيغة المجرَّدة Form من الزَّوائِد واللَّواصق؛ للكشف عن الجزء الأوَّل من السِّمة الكلاميَّة للقسم الكلاميِّ الَّذي تنتمي إليه الصِّيغة، ثم تتجلَّى السِّمة الأخرى بالبحث في الزَّوائِد؛ للكشف عن الفصائل النَّحوِيَّة، وهذا ما يبدو في نموذج مُفْرَدة «اللَّاسِلْكِيَّةُ»، إذ تجرَّدت الصِّيغة من الزَّوائد واللَّواصق، فأصبحت «لَاسِلْكيّ» التي استُدِلَّ عليها بالقسم الكلاميِّ A (الصفة)، ثُمَّ انطلق القسيم الثاني للسِّمة الكلاميَّة الأخرى من الزَّوائِد (أل التَّعرِيف، وتاء التَّأنيث، وعلامة الإعراب)؛ لتعبِّر عن الفصائل النَّحوِيَّة التي تحملها المفردة، فكانت الرُّموز FS1D دلالة على المؤَنَّث المفرد المرفوع المعرَّف.

(1) Smrž, O & Bielický, V. et al., (2008). Ibid.

(2) Smrž, O & Bielický, V. et al., (2008). Ibid.

Form	All~Asilokiy~apu	اللَّاسِلْكِيَّةُ
Morph	Al + lAsilokiy~ + ap + u	
Tag	DET+ADJ+NSUFF_FEM_SG+CASE_DEF_NOM	
Gloss	the + wireless / radio + [fem.sg.] + [def.nom.]	
Lemma	[lAsilokiy~_1]	لَاسِلْكِيّ
Root	*implicit in the lexicon*	

الشكل:٢١ مخرجات محلل **Buckwalter** –نموذج " اللَّاسِلْكِيَّةُ " [(١)]

أمَّا الفئة الكلاميَّة المستخدمة في محلِّل باكولتر الصَّرفِيِّ فقد تستند على الصَّرف الشَّكليِّ Morph الذي يتمثَّل في تحديد الوحدة الصَّرفيَّة أو المورفيم، إذ نجد في النَّموذج أعلاه أن السِّمة الكلاميَّة DET+ADJ+NSUFF_FEM_SG+CASE_DEF_NOM قد انطلقت من تفكيك الكلمة «اللَّاسلكية» إلى مورفيمات متتالية «ال+لاسلكي+ية+ُ»، ثم وَصْف كل مورفيم على حدة.

وقد اشتملت تلك الفئة الكلاميَّة على خمسة أقسامٍ رئيسيَّةٍ (الفعل – الاسم – الضمير – الأداة – أخرى) تتخلَّلها عدَّة أقسامٍ فرعيَّةٍ، لتصل بذلك إلى ٢٠ قسمًا، كما تضمنت عددًا من الفصائل النَّحوِيَّة، ليبلغ مجمل عدد سمات الفئة الكلاميَّة ٣٤ سمة يوضِّحها الجدول رقم ١٣:

المصطلح العربي	المصطلح الانجليزي	الاختصار
	الفعل Verb	
فعل مضارع	imperfect verb	VI
فعل ماض	perfect verb	VP
فعل أمر	imperative verb	VC

(1) Smrž, O & Bielický, V. et al., (2008). Ibid.

المصطلح العربي	المصطلح الانجليزي	الاختصار
الاسم Noun		
اسم	noun	N
اسم علم	proper noun	Z
صفة	adjective	A
حال	adverb	D
اختصار	abbreviation	Y
الضمير Pronoun		
ضمير	pronoun	S
اسم إشارة	demonstrative pronoun	SD
اسم موصول	relative pronoun	SR
الأداة Particle		
أداة	particle	F
أداة استفهام	interrogative particle	FI
أداة نفي	negative particle	FN
حرف عطف	conjunction	C
حرف جر	preposition	P
حرف تعجب	interjection	I

المصطلح العربي	المصطلح الانجليزي	الاختصار
أخرى Other		
رمز رسومي	graphical symbol	G
عدد	number	Q
أداة تعريف معزولة	isolated definite article	--

السمات Features	
الإعراب	INdicative (الرفع) - Subjunctive(النصب)- Jussive (الجزم)-D (في حالة الاشتباه بين حالة النصب والجزم)
صيغة البناء	Active(معلوم)- Passive (مجهول)
علامة الشخص	1(للمتكلم)-2 (للمخاطب)-3 (للآخرين)
علامة الجنس	Masculine(للمذكر)- Feminine (للمؤنث)
علامة العدد	Singular (المفرد)-Dual (المثنى)- Plural (الجمع)
الحالة الإعربية	1=nominative (للرفع)- 2=genitive (للنصب)-4=accusative (للجر)
التعريف	Indefinite (نكرة)-Definite (معرفة)-Reduced (مقلصة)-Complex (مركبة).

جدول:١٣ فئة بادت الكلامية (١)

تلتقي معظم سمات تلك الفئة الكلاميَّة مع العديد من الفئات الكلاميَّة الأخرى، غير أنَّ وضع بعض السِّمات تحت أقسامها الرئيسيَّة يختلف نوعًا ما من فئةٍ لأخرى، وهذا يعود إلى طبيعة العَنْوَنة التَّمثيليَّة لهذه السِّمات. ومع ذلك نلحظ تَميُّز هذه الفئة الكلاميَّة ببعض السِّمات الكلاميَّة مثل تصنيفها للتَّعرِيف كأداة Particle وفي الوقت ذاته فصيلة نحوِيَّة (معرِفة)، إلَّا أنَّ الآخر لا يكون قسيمًا إلا لأداة التَّعريف المستقِلَّة،

(1) Smrž, O. & Bielický, V. et al., (2008). Ibid.

بَيْنَما النَّوْعان الآخران لفصيلة التَّعرِيف: المقلَّصة و المُرَكَّبة مساويان لحالة البناء، غير أن المركَّبة تمتلك أداةً للتَّعرِيف مثل الإضافة غير الحقيقيَّة.[١]

٢-٥ فئة آر دي آي الكلامية RDI POS Tag set

في ظل إبراز الحاجة لمشكلٍ آليٍّ عند بناء نظم معالجة الكلام العربيِّ آليًّا وخاصة توليد الكلام المنطوق من النَّصِّ العربيِّ المكتوب، فقد استوجب التَّبصُّر في معطيات المستويات اللِّسانيَّة المختلفة بما يتوافق مع تقنيات الهندسة اللُّغويَّة لا سِيَّما تقنية الذكاء الاصطناعيِّ التي تتطلَّب توصيفًا لُغوِيًّا مطردًا نابعًا من استقراء النِّظام اللُّغويِّ، يتلاءم مع ما تتطلَّبه آليَّة التَّشكيل الآليِّ.

وفي سبيل ذلك قدَّم الدُّكتور محمد عطية[٢] الباحث بالشركة الهندسيَّة لتطوير النُّظم الرقميَّة بالقاهرة تصميمًا لفئةٍ كلاميَّةٍ ضمن أطروحته للدُّكتوراة، ينطلق من التَّحلِيل الصَّرفيِّ[٣]باعتباره دليلًا هاديًّا في التعرُّف على الأنواع الكلاميَّة، إذ يعتمد على المَكْنَز المورفيميِّ العربيِّ الذي يحتوى على ٧٨٠٠ مورفيم في استقراء الخصائص الصَّرف - نحويِّة لكل كلمة عربيَّة ممكنة.[٤]

وكانت نتيجة الاستقراء والاختصار الحصولَ على أصغر فئة كلاميَّة ممكنةٍ من المورفيمات العربيَّة، تحتوي على ٦٢ نوعًا كلاميًّا، يعرضها الجدول رقم ١٤:

---

(1) Habash, N & Roth, R. (2009). CATIB: The columbia Arabic Treebank.Proceedings of the ACL-IJCNLP 2009 Conference Short Papers.Suntec, Singapore.

(٢) يشغل عدة مناصب هامة منها:
استشاري معالجة اللغات الطبيعية بالشركة الهندسية لتطوير النظم الرقمية - القاهرة.
استشاري هندسة البرمجيات بشركة لوكسور تكنولوجي - كندا.
أستاذ زائر علوم حاسب وتكنولوجيا المعلومات بالأكاديمية العربية للعلوم والتكنولوجيا والنقل البحري - مصر.
محكم دائم في أكثر من مجلة علمية رفيعة المستوى.
عضو في لجنة إعداد المعجم التاريخي للغة العربية بالدوحة.

(٣) يتمثل في المحلل الصرفي «ميزان» الذي يعتمد على المورفيم كلبنة بنائية أساسية، وهو ما يميزه عن نظرائه مما يرتكز على حصيلة من الكلمات، وهو ما يعزز من مرونته وتغطيته لمفردات اللغة العربية والتي تتخطى نسبة ٩٩,٨٪.. للاطلاع والتجريب: RDI. (1993). http://www.rdi-eg.com/ar/Technologies/arabic_nlp.htm

(4) Attia, M. (2004).Ibid.P.43.

الفئة	المصطلح العربي	المصطلح الإنجليزي	الاختصار
علامة بداية الكلمة	بداية كلمة	Start-Of-Word marker	SOW
الحشو	حشو	Padding string	Padding
سمات سوابق الاسم والفعل	لا سابق	Null prefix	NullPrefix
	عطف	Conjunctive	Conj
	لام التوكيد	Confirmation by Laam	Confirm
	همزة استفهام	Interrogation by Hamza	Interrog
سمات لواحق الاسم والفعل	لا لاحق	Null suffix	NullSuffix
	ضمير نصب أو جر	Object or possession pronoun	ObjPossPro
سمات الحالات النحوية للفعل والاسم	مرفوع	1st Arabic syntactic case	MARF
	منصوب	2nd Arabic syntactic case	MANSS
سمات سوابق الاسم فقط	«ال» التعريف	Definitive article	Definit

الفئة	المصطلح العربي	المصطلح الإنجليزي	الاختصار
سمات جذع الاسم فقط	اسم	Nominal	Noun
	مصدر	Nouns made of infinitives	NounInfinit
	اسم مصدر	NounInfinit" like"	NounInfinitLike
	اسم فاعل	Subject noun	SubjNoun
	صيغة مبالغة	Exaggeration adjective	ExaggAdj
	اسم مفعول	Object noun	ObjNoun
	اسم زمان أو مكان	Noun of time or location	TimeLocNoun
	ممنوع من الضرف	An Arabic feature of a specific class of nouns	NoSARF
سمات لواحق الاسم فقط	ضمير جر	Possessive pronoun	PossessPro
	نسب	Relative adjectives maker	RelAdj
	تأنيث	Feminine	Femin
	مذكر	Masculine	Masc
	مفرد	Singular	Single
	مثنى	Binary	Binary
	جمع	Plural	Plural
	مضاف	Adjunct	Adjunct
	غير مضاف	NonAdjunct	NonAdjunct

الفئة	المصطلح العربي	المصطلح الإنجليزي	الاختصار
سمات لواحق الاسم فقط	منصوب أو مجرور	2nd or 3rd Arabic syntactic case	MANSS_MAGR
	مجرور	3rd Arabic syntactic case	MAGR
سمات سوابق الفعل فقط	مضارع	Present tense	Present
	استقبال	Future tense	Future
سمات جذع الفعل فقط	مَبْنِيٌّ للمعلوم (للفاعل)	Active sound	Active
	مَبْنِيٌّ للمجهول (للمفعول)	Passive sound	Passive
	أمر	Imperative	Imperative
	فعل	Verb	Verb
	لازم	Transitive verb	Transitive
	مجزوم	4th Arabic syntactic case	MAJZ
	ماض	Past tense	Past
	مضارع أو أمر	Present tense, or imperative	PresImperat
سمات لواحق الفعل فقط	ضمير رفع	Subject form pronoun	SubjPro
	ضمير نصب	Object form pronoun	ObjPro
	منصوب أو مجزوم	2nd or 4th Arabic syntactic case	MANS_MAJZ

الفئة	المصطلح العربي	المصطلح الإنجليزي	الاختصار
سمات وظيفية ثابتة	حرف جر	Preposition	Prepos
	حرف نداء	Interjection	Interj
	جار ومجرور	Preposition-Pronoun Compound	PrepPronComp
	اسم موصول	Relative pronoun	RelPro
	اسم إشارة	Demonstrative pronoun	DemoPro
	أداة استفهام	Interrogation article	InterrogArticle
	جازمة	For specific articles that make the consequent verb in the 4th Arabic syntactic case	JAAZIMA
	شرطية جازمة	Feature of a class of Arabic conditionals	CondJAAZIMA
	شرطية غير جازمة	Feature of a class of Arabic conditionals	CondNot JAAZIMA
	لا	Arabic specific article	LAA
	لات	Arabic specific article	LAATA
	استثناء	Article of exception	Except
	غير عاملة	A class of articles that have no syntactic effect	NoSyntaEffect
	ظرف	Feature for certain kind of Arabic adverbs	DZARF

الفئة	المصطلح العربي	المصطلح الإنجليزي	الاختصار
سمات وظيفية ثابتة	حرف ناسخ	A class of particles that make the subject of the consequent nominal sentence in 2nd Arabic syntactic case	ParticleNAASIKH
	فعل ناسخ	A class of auxiliary verbs that make the predicate of the consequent verbal sentence in 2nd Arabic syntactic case	VerbNAASIKH
	ناصب	Arabic specific class of particles that make the consequent verb in 2nd Arabic syntactic case	ParticleNAASSIB
	مصدرية	Arabic specific article	MASSDARIYYA
الكلمة الأجنبية	كلمة أجنبية مكتوبة بحروف عربية	Transliterated Arabic string	Translit

جدول: ١٤ فئة آر دي آي الكلامية

وبِمُلاحظة فئة الأنواع الكلاميَّة في الجدول السابق، يتبيّن ما يلي:

- انطلاق الفئة الكلاميَّة من العنصر التَّحليليِّ المورفيميِّ بعكس ما تسير عليه الفئات الأخرى التي كان منطلقها القسم الكلاميِّ لا المورفيم؛ لذا تُسمّي كل مورفيم بالنوع الكلاميِّ.

• تشتمل الفئة الكلاميَّة على عنصرَيْنِ غيرِ لُغَوِيَّيْنِ، وهما بداية الكلمة SOW والحشو[1] Padding؛ للتَّوَافُق مع أسلوب التَّعلُّم الإحصائيِّ المنشود.

• تتفرَّد الفئة الكلاميَّة بالعديد من المورفيمات العربيَّة التي تختصّ بجذع الكلمة ولَواحقها.

• تطرَّقت الفئة الكلاميَّة لسمةٍ نحوِيَّةٍ سياقِيَّةٍ ألَا وهِيَ الجارُّ والمجرور PrepPronComp لاستخدامها مع حرف الجرِّ – ذي المحرف الواحد- المتشابك مع الضَّمير، مثل (بها، لها به، له، ...).

• تُؤرجِح الفئة الكلاميَّة بين الحالات النَّحويَّة للمفردات باستخدام السِّمة القطعيَّة مثل «منصوب، مجرور» والسِّمة الاحتماليَّة «منصوب أو مجرور، منصوب أو مجزوم».

• تعبِّر الفئة الكلاميَّة عن الكلمة الأجنبيَّة بكل ما لا يُدْرَج في مظانَّة النَّموذج الصَّرفيِّ Morphological Model الذي بنيت عليه هذه الفئة.

وقد اعتمد الدُّكتور عطية على منهجية متكاملة في التَّصميم والاستقراء والاختصار،[2] يمكننا أن نلخصها كالآتي:

• الشُّموليَّة، بحيث تشتمل الفئة الكلاميَّة على أي سمة نحويَّةٍ غير سياقيَّة في أي كلمة عربيَّة.

• الاطراد والتَّناسُق، بحيث لا يوجد في الفئة الكلاميَّة سمة أو نوع كلاميٍّ يكافئه نوع كلاميٌّ آخر.

• الانتقاء والإيجاز، بحيث يتمُّ استقراء المكْنَز المورفيميِّ، ومن ثَمَّ استخلاص أصغر فئة كلاميَّة ممكنة.

وبناءً على توخِّيه هذه المبادئ التي انتهجها في تصميم فئته الكلاميَّة، فقد قام بتسميَة

(١) أي حشو المتجه الكلامي بتلك السمة، بحيث يتم التوحيد في عدد الأنواع الكلاميَّة لكل كلمة؛ حتى تتوافق كل الكلمات في عدد الأنواع الكلاميَّة؛ تلبية لآلية التعلُّم الإحصائيِّ.

(2) Attia, M. (2004).Ibid.P.44.

كل مورفيم في المكْنَز المورفيميِّ بالأنواع الكلاميَّة التي تستوفي خصائصه النَّحوِيَّة[١]، ويوضِّح الجدول رقم ١٥ أمثلة على تسمية المورفيمات العربيَّة بالأنواع الكلاميَّة.

المورفيم العربي	نوعه في السلسلة النصية	متجه العناوين الكلامية
سابقة؛ ٩	الـــ	[ال التعريف]
سابقة؛ ١٢٥	سَيـــ	[استقبال، مضارع، مبني للمعلوم]
صيغة اشتقاقية منتظمة؛ ٤٨٢	مُفَاعِل	[اسم، اسم فاعل]
صيغة اشتقاقية منتظمة؛ ٦٧	اسْتِفْعَال	[اسم، مصدر]
صيغة اشتقاقية غير منتظمة؛ ٢٩	مَلائِك	[اسم، ممنوع من الصرف، جمع]
جامد؛ ٨	هُوَ	[اسم، مذكر، مفرد، ضمير رفع]
جامد؛ ٣٩	ذُو	[اسم، مذكر، مفرد، مضاف، مرفوع]
لاحقة؛ ٢٧	ـــات	[مؤنث، جمع]
لاحقة؛ ٤٢٧	ـــونَهُمْ	[مضارع، مرفوع، ضمير رفع، ضمير نصب]
لاحقة؛ ١٩٥	ـــيَّتَان	[نسب، مؤنث، مثنى، غير مضاف، مرفوع]

الجدول:١٥ نموذج تسمية بعض المورفيمات بالأنواع الكلامية [٢]

اعتمد الدُّكتور عطية في بناء عمليَّة العَنْوَنة الكلاميَّة على التَّحلِيل الصَّرفيِّ، حيث اتَّخذ مورفيم الصِّيغة – من حيث كونه مصرَّفًا أو غير مصرَّف- دليلًا مستأنسًا في تحديد

(1) Attia, M. (2004).Ibid. P.46.

(2) Attia, M. (2004).Ibid. P.46.

خصائص جذع الكلمة، وكذلك مورفيمات السَّوابق واللَّواحق في تحديد خصائص لواصق الكلمة وزوائدها. وهو بذلك يُغاير الطُّرُق المعهودة في بناء عمليَّة العَنْوَنة؛ لأنَّه يوظّف العَنْوَنة في استنباط علامات التَّشكيل الإعرابيَّة للكلمات السَّابق تحليلها صرفيًّا. ويعرض الجدول رقم ١٦ مثالًا مُعَنْوَنًا بالأنواع الكلاميَّة (اللُّغويَّة) – التي تظهر في شكل متَّجهات – لجملة «وقد صرحت رئيسة الوزراء في نيوزيلندا».

الكلمة	متَّجه العناوين الكلامية	POS Tags vector
وَقَدْ	[عطف، غير عاملة، لا لاحق]	[Conj,NoSyntaEffect,NullSuffix]
صَرَّحَتْ	[لا سابق، فعل، ماض، مفرد، مؤنث]	NullPrefix,Verb,Past,Single,Fe] [min
رَئِيسَة	[لا سابق، اسم، مفرد، مؤنث]	[NullPrefix,Noun ,Single,Femin]
اَلْوُزَرَاء	[ال التعريف، اسم، جمع، ممنوع من الصرف، لا لاحق]	Definit,Noun,Plural,NoSARF,N] [ullSuffix
فِي	[لا سابق، حرف جر، لا لاحق]	[NullPrefix,Prepos,NullSuffix]
نيوزيلندا	[كلمة أجنبية مكتوبة بحروف عربية]	[Translit]

الجدول:١٦ نموذج معنون بفئة آر دي آي الكلامية [١]

٢-٦ الفئة الكلامية المقلصلة الموسعة

(The Extend Reduced Tag set ERTS)

أدركت الدُّكتورة منى دياب مدى القصور الذي لحقته فئة بييز الكلاميَّة بعد أنْ عجزت عن مُقاربَة احتواء اللُّغة العربيَّة، فآثرت أنْ تُضيف بعض السِّمات التي تتفق ووجهةَ نظرها من حيث ملاءمتها لطبيعة اللُّغة العربيَّة، وذلك لاستخدامها في نظام

(1) Attia, M. (2004).Ibid. P.47-48.

أميرة AMIRA وهو نظام متعدِّد المهامِّ (تجزئة النُّصوص، العَنْوَنة بالأقسام الكلامِيَّة، التَّحلِيل النَّحوِيِّ السَّطحِيِّ Shallow Syntactic Parsing) يرعاهُ مركز أنظمة التَّعلُّم الحاسوبِيِّ CCLS بجامعة كولومبيا بالولايات المُتَّحِدَة الأمريكيَّة. [١]

وقد اقتبست فئتها الكلاميَّة من فئة باكولتر الكاملة لِلنَّصِّ المجزَّأ؛ لتبلغ بالإضافة لفئة بييز ٧٢ سمة، وتمثَّلت الإضافات في الفصائل النَّحوِيَّة (العدد والجنس، والتَّعرِيف)، إلا أنها اقتصرت على فئة الأسماء والصفات فحسب، فَتَرْمُز للعدد بالرَّمز DU للمثنَّى، و بالرَّمز S للجمع، بَيْنَما تُرمِّز للجنس بالرَّمز F للمُؤَنَّث و بالرَّمز M للمُذَكَّر، أمَّا التَّعرِيف (وتحديدًا أداة التَّعرِيف) فَتَرْمُز أداة التَّعرِيف التي تظهر مع الأسماء بالرَّمز D. [٢]

وقد قارنت الدُّكتورة منى دياب بين فئة باكولتر الكاملة وبين الفئتَيْن الكلاميَّتَيْن المستخدمتَيْن في تحليل بنك بنسلفانيا العربيِّ - اللَّتَيْن قدمتهما مؤسسة شبكة البيانات اللُّغوِيَّة LDC – من خلال جملة «حصيلة نهائية حادث النار الجماعي شخصين»؛ لتبيّن أوجه الاختلاف بين الفئات الكلاميَّة الثلاث كما في الجدول رقم ١٧.

الكلمة	فئة باكولتر الكاملة FULL	فئة بييز RTS	الفئة المقلصلة الموسعة ERTS
حصيلة	NOUN+ NSUFF_FEM_SG+CASE_IND_NOM	NN	NNF
نهائية	ADJ+ NSUFF_FEM_SG+ CASE_IND_NOM	JJ	JJF
حادث	NOUN+ CASE_DEF_ACC	NN	NNM
النار	DET+ NOUN+ CASE_DEF_GEN	NN	DNNM
الجماعي	DET+ ADJ+ CASE_DEF_GEN	JJ	DJJM

(1) Diab, M. (2007). Improved Arabic Base Phrase Chunking With a new enriched POS tag set. Proceedings of the 5th Workshop on Important Unresolved Matters, Prague, Czech Republic.

(2) Habash , N. (2010). Ibid. P.83.

الكلمة	فئة باكولتر الكاملة FULL	فئة بييز RTS	الفئة المقلصة الموسعة ERTS
شخصين	NOUN+ NSUFF_MASC_ DU_GEN	NNS	NNMDu

جدول:١٧ مقارنة بين الفئات الكلامية الثلاث(١)

ويتَّضح من عرض هذه الفئات الكلاميَّة أن الفئة المقلَّصة المُوسَّعة أكثر إحكامًا في حمل مورفيمات الفصائل النَّحوِيَّة، إذ نجد أن فئة الأسماء والصفات تحملانِ سمةَ التذكير في حال الإفراد بعكس ما تسير عليه فئة باكولتر، في حين تَتفرَّد فئة باكولتر بالعلامات الإعرابيَّة التي تحملها أجزاؤها الكلاميَّة، كما تعبّر عن مورفيمات الكلمة بمتَّجَهٍ كلاميٍّ من الرُّموز تحكمه العلامة الدالة على الجمع «+» ، بَيْنَما تعبّر الفئات الأخرى عن مورفيمات الكلمة برمزٍ كلاميٍّ واحدٍ.

٢-٧ فئة القريني الكلامية Al Qrainy tag set

قدّم الدُّكتور شحادة القريني ضمن أطرُوحته للدُّكتوراة فئةً كلاميَّةً لِلُّغة العربيَّة تُسمَّى ARBTAGS بغرض بناء مُعَنْونٍ كلاميٍّ Arabic Morphosyntactic Tagger (AMT) يستند على أكثر السِّمات الصَّرف- نحويَّة شيوعًا التي تشترك في مطلبها كافَّة التَّطبيقات اللُّغوِيَّة،(٢) وقد اعتمد القريني في تقسيمه للأقسام الكلاميَّة على التَّقسيم الثُّلاثيِّ (اسم – فعل –أداة) بالإضافة لعلامة التَّرقيم؛ ليصل بتقسيمه إلى أربعة أقسامٍ رئيسيَّةٍ تتخللها عدة أقسام فرعية، ثم أضاف في فئته تلك مورفيمات الفصائل النحوية والعلامات الإعرابية التي تحملها فئات الأفعال والأسماء.(٣)

وبهذا قد اشتملت تلك الفئة على ٢٨ قسمًا فرعيًّا من الأقسام الكلاميَّة، و١٧ مورفيمًا من الفصائل النَّحوِيَّة والعلامات الإعرابيَّة، وعند اتِّحاد هذه الأقسام الفرعيَّة مع مورفيمات الفصائل النَّحوِيَّة والعلامات الإعرابيَّة تصل الفئة الكلاميَّة إلى ١٦١ سمة،

(1) Diab, M. (2007). Ibid. P.92.

(2) Alqrainy, Sh. (2008). A Morphological - Syntaical Analysis Approach For Arabic Textual Tagging. Thesis PhD in Computer Science. De Montfort University. P.25.

(3) Alqrainy, Sh. (2008). Ibid.P.32.

تتوزَّع على الأقسام الرئيسيَّة كالآتي: ١٠١ اسم، ٥٠ فعلا، ٩ أدوات، علامة ترقيم، مثال ذلك كلمة «يشاهدون» تُعَنْون بـVePiMaP1ThDc (فعل مضارع_مذكر_جمع_ضمير الغائبين_مرفوع).(١)

ويوضِّح الجدول رقم ١٨ الفئة الكلاميَّة دون الفصائل النَّحوِيَّة والعلامات الإعرابية، فاشتمل الفعل على ثلاثة أقسام فرعيَّةٍ، والاسم على سبعة عشر قسمًا فرعيًّا، والأداة على سبعة أقسامٍ فرعيَّةٍ، بينما اشتملت علامة التَّرقيم على قسمٍ فرعيٍّ واحدٍ.

المصطلح العربي	المصطلح الانجليزي	الاختصار
الفعل		
فعل ماض	Perfect verb	VePe
فعل مضارع	Imperfect verb	VePi
فعل أمر	Imperative verb	VePm
الاسم		
اسم علم	Proper noun	NuPo
اسم شائع	Common noun	NuCn
اسم الصفة	Adjective noun	NuAj
اسم مصدر	Infinitive noun	NuIf
اسم منسوب	Relative noun	NuRe
اسم تصغير	Diminutive noun	NuDm
اسم آلة	Instrument noun	NuIs

(1) Alqrainy, Sh. (2008).Ibid.P36.

الاختصار	المصطلح الانجليزي	المصطلح العربي
NuPn	Noun of Place	اسم المكان
NuTn	Noun of Time	اسم الزمان
NuPs	Pronoun	الضمير
NuCv	Conjunctive noun	اسم وصل
NuCd	Conditional noun	اسم الشرط
NuDe	Demonstrative noun	اسم الإشارة
NuIn	Interrogrative noun	اسم استفهام
NuAd	Adverb	الظرف أو الحال
NuNn	Numeral noun	اسم عددي
Fw	Foreign noun	اسم أجنبي
	أداة	
PrPp	Preposition	حرف الجر
PrVo	Vocative Particle	أداة النداء
PrCo	Conjunction Particle	أداة عطف
PrEx	Exception Particle	أداة استثناء
PrAn	Annulment Particle	أداة إلغاء العمل
PrSb	Subjunctive Particle	أداة شرط

المصطلح العربي	المصطلح الانجليزي	الاختصار
أداة جزم	Jussive Particle	PrJs
	الترقيم	
علامة الترقيم	Punctuation mark	Pun

جدول:١٨ فئة القريني الكلامية[١]

ويلاحظ مدى التَّقارُب بين تلك الفئة وفئة خوجة الكلامية، حيث إنَّ كليهما ينطلق من النَّحو العربيِّ التَّقليديِّ إلَّا أنَّ فئة خوجة تُوَسِّع دائرة الأقسام الكلاميَّة، فتقسمها إلى خمسة أقسام رئيسيَّةٍ، بَيْنَمَا تزيد الأقسام الفرعيَّة في فئة القرينيِّ؛ لتُغَطِّي السِّمات الصَّرف – نحوِيَّة الأكثر استخدامًا في تطبيقات معالجة اللُّغة الطبيعيَّة.

٢–٨ فئة كاتب الكلامية The CATIB Tag set

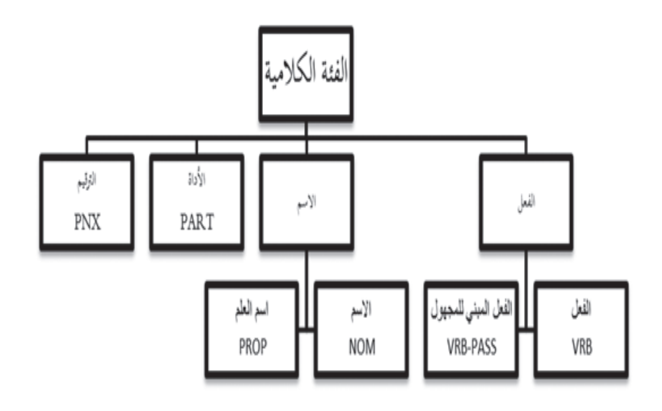
شرع مركز أنظمة التَّعلُّم الحاسوبيِّ CCLS بجامعة كولومبيا في عام ٢٠٠٨ م في بناء بنكٍ شجريٍّ لِلُّغة العربيَّة The Columbia Arabic Treebank (CATiB) بِدَعْمٍ من وزارة الدِّفاع الأمريكيَّة، بهدف التَّرجمة الآليَّة، وقد راعى في خُطَّته المنهجية تقليل الوقت والجهد اللَّذَيْنِ يخضعان لحجم المعلومات اللُّغوِيَّة المراد إضافتها لِلمادَّة الخامِّ؛ بغيةَ الإسراع من العَنْوَنة اليَدوِيَّة مع مراعاة المحافظة على الفروقات الهامَّة[٢]...وعليه فقد اتَّبع النهج المقتصد في إحكام حجم الفئة الكلاميَّة، حيث احتوت الفئة على سِتَّة أقسام كلاميَّةٍ.

وتَجْدُر الإشارة إلى أنَّ التَّقليل مع الإحكام في سمات الفئة الكلاميَّة قد أثبتا فائدتهما في ضوء استخدام أساليب التَّعلُّم الإحصائيِّ التي تكون قادرة على التَّنبُّؤ بدِقَّة كلما كانت معطياتها التَّوصيفيَّة أكثر تحديدًا وإحكامًا.

(1) Alqrainy, Sh. (2008).Ibid.P.33.

(2) Habash , N. (2010). Ibid. P.83.

وقد انطلقت الفئة الكلاميَّة من النَّحو العربيِّ التَّقليديِّ، فاعتمدتْ على التَّقسيم الثّلاثي مضيفةً إليه قسمًا رابعًا ألَا وهُوَ علامة التَّرقيم، ثم قسَّمت كل قسم من القسمَيْنِ الرئيسيَّيْنِ (الفعل والاسم) إلى أقسامٍ فرعيَّةٍ[١]...وبهذا تصل إلى ستةِ أقسامٍ يوضِّحها الشَّكل رقم ٢٢:



الشكل:٢٢ فئة كاتب الكلامية CATIB POS Tag set

على الرُّغْم من دِقَّة النَّتائج التي أحرزتها تلك الفئة الكلاميَّة في تطبيقها على بعض تطبيقات معالجة اللُّغة الطبيعيَّة، فإنها تفقد الكثير من السِّمات التي تُعين على تحديد العَلاقات التَّركيبيَّة في الجملة العربيَّة، ولذلك تمَّ توسعتها لاحقًا بربط السِّمة الكلاميَّة بالسَّابقة واللَّاحقة التي تشغلها في المفردة مثل كلمة «الكاتبون» التي كانت تُعَنْوِن بـ (NOM)، أصبحت تعنون هكذا (AL+NOM+wn) في ضوء التَّوسعة.[٢]

ويوضِّح الجدول رقم ١٩ مثالًا من مُدَوَّنة بنك كولومبيا الشجريِّ لجملة «وقال رود لارسن متحدثا للإذاعة العامة النرويجية « أن أر كاي» أنه« يمكن الآن تشبيه المنطقة من عدة نواح ببرميل بارود بفتيل مشتعل». موصَّفة بالأقسام الكلاميَّة.

(1) Habash , N. (2010). Ibid. P.83.

(2) Habash , N. (2010). Ibid. P.83.

الكلمة	POS Tags	الأقسام الكلامية
و+	PRT	أداة
قال	VRB	فعل
رود	PROP	اسم علم
لارسن	PROP	اسم علم
متحدثا	NOM	اسم
ل+	PRT	أداة
الإذاعة	PROP	اسم علم
العامة	PROP	اسم علم
النروجية	PROP	اسم علم
"	PNX	علامة ترقيم
أن	PROP	اسم علم
أر	PROP	اسم علم
كاي	PROP	اسم علم
"	PNX	علامة ترقيم
أن	PRT	أداة
هـ+	NOM	اسم
"	PNX	علامة ترقيم

الكلمة	POS Tags	الأقسام الكلامية
يمكن	VRB	فعل
الآن	NOM	اسم
تشبيه	NOM	اسم
المنطقة	NOM	اسم
من	PRT	أداة
عدة	NOM	اسم
نواح	NOM	اسم
ب+	PRT	أداة
برميل	NOM	اسم
بارود	NOM	اسم
+ب	PRT	أداة
فتيل	NOM	اسم
مشتعل	NOM	اسم
"	PNX	علامة ترقيم
.	PNX	علامة ترقيم

الجدول:١٩ فئة كاتب الكلامية – نموذج العنونة بالأقسام الكلامية [١]

(١) هذا المثال مقتبس من بنك كولومبيا الشجري، وقد زودني به الأستاذ الدكتور نزار حبش.

وبمُلاحظة المثال السابق يتبيَّن لنا الآتي:

• أنَّ عَنْوَنة هذه الفئة الكلاميَّة تعتمد بصورة سافرة على النَّصِّ المجزَّأ للزَّوائد Clitics.

• أن سمة NOM تشغل الاسم والصِّفة والظَّرف والضَّمير بنوعيه المتَّصل والمنفصل...، بينما سمة PROP تشغل أسماء الأعلام المفردة والمركَّبة.

• أن سمة PRT تشغل حروف العطف والجرّ والحروف المصدرِيَّة وأدوات الرَّبط.

• أن سمة VRB تشغل الفعل المبني للمعلوم بالإضافة إلى أفعال كان وأخواتها وكاد وأخواتها، بينما سِمة VRB-pass تقتصر على الفعل المبني للمجهول.

٢-٩ فئة كاليك الكلامية The Kulick Tag set

قام سيث كاليك Seth Kulick في إطار المشروع البحثيِّ الذي ترعاهُ مُؤَسَّسة شبكة البيانات اللُّغوية LDC للمقارنة بين دِقَّة التَّحليلات النَّحوِيَّة الآليَّة لِلُّغات المختلفة لا سِيَّما بين العربيَّة والإنجليزيَّة وكذلك الصينيَّة بتصميم فئةٍ كلاميَّةٍ لِلُّغة العربيَّة تماهي ما يرتكز عليه محلِّل بايكل Bikel's parser لِلُّغة الإنجليزيَّة من فئة بنك بنسلفانيا الكلاميَّة باعتباره معيارًا ذهبيًّا للمقارنة بين تلك النَّماذج النَّحويَّة المختلفة؛ وذلك لكفاءته التَّحلِيليَّة.(١)

وسعيًا وراء بُلوغ الهدف المنشود، فقد اعتمد كاليك في فئته الكلاميَّة على فئة بييز مُضيفًا إليها بعض السِّمات؛ كي تتناسبَ والنموذجَ المُقارن أي فئة بنك بنسلفانيا، ويمكن أن تصنَّف هذه السِّمات إلى خمسة تصنيفاتٍ:(٢)

• تُرمّز علامات التَّرقيم برمز يتوافق نسبيًّا مع شكلها الفعليِّ، إذ ترِد في الفئة الإنجليزيَّة بهذا النَّمط، فعلامات التَّرقيم التي تعبِّر عن نهاية جملة مثل النُّقطة (.)، الفاصلة (؟)، علامة التعجب (!) تُرمَّز جميعها بنقطة؛ لتصبح كالآتي: (. .)، (؟.)،

(1) Kulick, S. (2010). Simultaneous Tokenization and Part-of-Speech Tagging for Arabic without a Morphological Analyzer. Proceedings of the ACL 2010 Conference Short Papers. Uppsala, Sweden.

(2) Habash , N. (2010). Ibid. P.82.

(! .)، أما علامات التَّرقيم النقطتان الرأسيَّتان (:)، الشَّرطة(–)، الفاصلة المنقوطة (؛) تُرَمَّز جميعها بنقطتين رأسيَّتَيْن؛ لتصبح كالآتي: (: :)، (– :)، (؛ :)، بَيْنَما علامات التَّرقيم الأُخرى تُرَمَّز بالعلامة نفسها مثل الفاصلة (، ،)، علامة التَّنصيص ("" ")، القَوْس اليمين(( ))، القوس الشمال(( )).

- إضافة الفئة الفرعيَّة (اسم عدد/ كمية NOUN_QUANT) في قسم الأسماء.

- إضافة الفئات الفرعيَّة (الصِّفة المقارنة ADJ_COMP) و(الأعداد التَّرتيبيَّة الوصفيَّة ADJ_NUM) في قسم الصِّفات.

- تمييز أداة الإشارة والتَّعريف بالرَّمْزَيْنِ (DEM) و(DT) على التَّوالي، حيث كانا يُرَمَّزانِ في فئة بييز برمز واحد DT وفي حالة الفصل فقط.

- ومن ثَمَّ ظهور علامة DT مجاورة لبعض السِّمات المركَّبة مثال ذلك: ال+اسم DT+NN و ال+صفة مقارنة ADJ_COMP+DT و ال+ عدد DT+CD، و ال+صفة DT+JJ.

وبهذا يصل كاليك بعدد السِّمات التي انتهجها بالإضافة لفئة بييز إلى ٤٣ سمة، وقد استخدمتْ في عَنْوَنة بنك بنسلفانيا العربيِّ الإصدار الثالث بهدف المقارنة التي يتم بِمُوجبها اكتشاف مواضع التَّقصير وسُبُل المعالجة.[١]

٢–١٠ الفئة الكلامية للنص القرآني The Quranic Text Tag set

بدأ الاعتناء بقضايا معالجة اللُّغة العربيَّة آليًّا في مطلع القرن الحادي والعشرين؛ لُمحاولة اقتصاد المعرفة في ظلِّ الانفجار المعلوماتيِّ الذي أدَّى بدوره لاتِّساع الفجوة المعرفيَّة؛ نتيجة لظهور الفجوة اللُّغوِيَّة التي ساهمت في تخلُّفٍ لُغويٍّ عن ركب التَّقدم التُّكنولوجِيِّ، فتولى فريق معالجة اللُّغات الطبيعيَّة بجامعة ليدز Leeds مشروعات متعدِّدة في بناء المدوَّنات اللُّغوية للنَّصِّ العربيِّ وتطوير آليَّات التَّحلِيل الصَّرفيِّ والتَّجذيع، وغيرها من متطلَّبات المعالجة الآليَّة لِلُّغة الطبيعيَّة، ثم امتدَّت تلك المشروعات لتشمل

(1) Habash , N. (2010). Ibid. P.82.

المعالجة اللُّغوية للقرآن الكريم التي أسهمت في تطوير آليَّات التَّنقيب في نصوص القرآن الكريم، مثال ذلك موقع «قرآني» الذي يتيح البحث في حيثيات المفاهيم القرآنيَّة.[١]

وقد لَاقت فكرة الموقع القرآني نجاحًا في تطبيقها، الأمر الذي أدّى إلى التَّفكير في تدشين مدوَّنةٍ لُغوِيَّةٍ للقرآن الكريم على الشَّبكة العنكبوتيَّة تتضمن معلوماتٍ لغويةً على المستوى الصَّرفِيِّ والأجزاءِ الكلاميَّة لِلنُّصوص القرآنيَّة[٢]، وقد انطلقت المدوَّنة القرآنيَّة في صورة تمثيلٍ معرِفِيٍّ لألفاظ القرآن الكريم في إطار هذه المعلومات اللُّغوية التي تُقَدِّمها؛ حتى تستقطب العديد مِمَّن يَنْشِدُونَ تعلُّم لُغة القرآن الكريم من غير النَّاطقين بها، وكذلك الباحثين في علوم اللُّغة العربيَّة؛ لِلبحث في معطيات التَّوصيف اللُّغوي لِلنَّصِّ القرآنيِّ.

وقد ارتكزت المعلومات اللُّغوِيَّة وخاصَّة الأقسام الكلاميَّة المتضمنة في تلك المدوَّنة القرآنيَّة على النَّحو التَّقليديِّ -لارتباطه بالنَّصِّ القرآنيِّ- حيث ارتبطت بالمصادر النَّحوِيَّة القديمة التي أُخِذَ منها أصوله ومصطلحاته، فكان ذلك أمرًا له بالغ الأثر، إذ ساعد اللِّسانِيِّين على عَنْوَنة المحتوى القرآنيِّ بسهولة ويسر لما تتيحه هذه المصادر من شواهِدَ وإعرابٍ لنصوص القرآن الكريم.

ولما كانت الغاية من الفئة الكلاميَّة أن تراعي أدقَّ التَّفصيلات النَّحوِيَّة؛ لتتناسب مع طبيعة النَّصِّ القرآنيِّ فقد اقتضى ذلك أن يُقَسِّموا الأقسام الكلاميَّة إلي ثلاثة أقسامٍ رئيسيَّةٍ تتضمَّنها تسعة أقسامٍ فرعيَّةٍ، ثُمَّ جعلوا تحت كلِّ قسمٍ فرعيٍّ أقسامًا أخرى... وبهذا تشتمل الفئة الكلاميَّة على ٤١ سمة[٣]، يوضِّحُها الجدول رقم ٢٠:

(1) Sharaf, A. & Atwell, E. & Dukes, K. and Others. (2010). Arabic and Quranic computational linguistics projects at the University of Leeds. In Proceedings of the workshop of Increasing Arabic Contents on the Web, organized by Arab League Educational, Cultural and Scientific Organization (ALECSO).

(2) Sharaf, A. & Atwell, E. & Dukes, K. and Others. (2010).Ibid.

(3) Dukes, K. & Buckwalter, T. (2010). A Dependency Treebank of the Quran using Traditional Arabic Grammar. 7th international conference on Informatics and Systems. Cairo, Egypt.

الأقسام الرئيسية	المصطلح العربي	المصطلح الإنجليزي	الاختصار
الاسم	الاسم		
	اسم	Noun	N
	اسم علم	Proper noun	PN
	اسم فعل أمر	Imperative verbal noun	IMPN
	الضمير		
	ضمير	Personal pronoun	PRON
	اسم إشارة	Demonstative pronoun	DEM
	اسم موصول	Relative pronoun	REL
	الصفة		
	صفة	Adjective	ADJ
	رقم	Number	NUM
الاسم	الظرف		
	ظرف زمان	Time adverb	T
	ظرف مكان	Location adverb	LOC
الفعل	فعل		
	فعل	Verb	V
الأداة		حرف	
	حرف جر	Preposition	P
	اللام –السوابق		
	لام التوكيد	Emphatic lâm prefix	EMPH
	لام الأمر	Imperative lâm prefix	IMPV
	لام التعليل	Purpose lâm prefix	PRP

الأقسام الرئيسية	المصطلح العربي	المصطلح الإنجليزي	الاختصار
		العطف	
	حرف العطف	Coordinating conjunction	CONJ
	حرف مصدري	Subordinating conjunction	SUB
		الأدوات	
	أداة نصب	Accusative particle	ACC
الأداة	أداة استدراك	Amendment particle	AMD
	أداة جواب	Answer particle	ANS
	أداة ردع	Aversion particle	AVR
	أداة سببية	Particle of cause	CAUS
	أداة تحقيق	Particle of certainty	CERT
	أداة شرط	Conditional particle	COND
	أداة تسوية	Equalization particle	EQ
	أداة تحضيض	Exhortation particle	EXH
الأداة	أداة تفصيل	Exceptive particle	EXL
	أداة استثناء	Exceptive particle	EXP
	أداة استقبال	Future particle	FUT
	حرف ابتداء	Inceptive particle	INC
	حرف استفهام	Interrogative particle	INEG
	حرف نفي	Negative particle	NEG
	حرف كاف	Preventive particle	PREV
	حرف نهي	Prohibition particle	PRO
	حرف استئناف	Resumption particle	REM
	أداة حصر	Restriction particle	RES
	حرف إضراب	Retraction particle	RET
	حرف زائد	Supplemental particle	SUP

الأقسام الرئيسية	المصطلح العربي	المصطلح الإنجليزي	الاختصار
	حرف فجأة	Surprise particle	SUR
	حرف نداء	Vocative particle	VOC
		حروف مقطعة	
	حروف مقطعة	Quranic initials	INL

الجدول: ٢٠ الفئة الكلامية للنص القرآني [١]

يلاحظ أن الأدوات قد نالت النَّصيب الأكبر في الفئة الكلاميَّة؛ لتعدُّد معانيها في القرآن الكريم، ولأهميَّتها في سياق النَّصِّ القرآنيِّ بوصفها مكونًا أساسيًّا في تحديد دلالة التَّركيب القرآنيِّ، وفي المقابل نجد الفعل أقلَّ حظًّا في تلك الفئة الكلاميَّة، حيث لم تتعرَّض لأنواع الفعل من حيث دلالاته على الزَّمن أو الحدث، إلا أنَّهم لم يغفلوا – أي فريق معالجة اللغة الطبيعية – عن ذلك، فقد أوضحوا تقسيمات الفعل والفصائل النَّحوِيَّة التي تحملها الأقسام الكلاميَّة في التَّحلِيل التَّفصيليِّ للكلمة بعيدًا عن التَّحليل المجمل لِلجُمَل القرآنيَّة كما ورد في الجدول السابق، ومن ثَمَّ لم يذكروا تلك التَّفصيلات في الفئة الكلاميَّة.

ويعرِض الشَّكل رقم ٢٣ الأقسام الكلاميَّة لِلْجُملة القرآنية «وَلَقَدِ اسْتُهْزِئَ بِرُسُلٍ مِّن قَبْلِكَ» المقتبسة من المدوَّنة القرآنيَّة التي أتاحتها جامعة ليدز البريطانيَّة على الشَّبكة العنكبوتيَّة.

Translation	Arabic word	Syntax and morphology
(6:10:1) walaqadi And indeed	وَلَقَدِ CERT EMPH REM	REM – prefixed resumption particle EMPH – emphatic prefix lām CERT – particle of certainty الواو استئنافية اللام لام التوكيد حرف تحقيق
(6:10:2) us'tuh'zi-a were mocked	ٱسْتُهْزِئَ V	V – 3rd person masculine singular (form X) passive perfect verb فعل ماض مبني للمجهول
(6:10:3) birusulin Messengers	بِرُسُلٍ N P	P – prefixed preposition bi N – genitive masculine plural indefinite noun جار ومجرور
(6:10:4) min from	مِّن P	P – preposition حرف جر
(6:10:5) qablika before you	قَبْلِكَ PRON N	N – genitive noun PRON – 2nd person masculine singular possessive pronoun اسم مجرور والكاف ضمير متصل في محل جر بالاضافة

الشكل:٢٣ جملة قرآنية معنونة بالأقسام الكلامية–سورة الأنعام آية ٦ [٢]

(1) Dukes, K. & Buckwalter, T. (2010).Ibid.

(2) Quran Syntax. (2009). http://corpus.quran.com/treebank.jsp

ويتَّضح من عَنْوَنة الجملة القرآنيَّة السابقة أنَّ تلك العَنْوَنة تَتَوَخَّى الطُّرق البصرِيَّة التَّمثيلِيَّة في تحديد الأقسام الكلاميَّة، علاوة على تعيين كلِ قسم كلاميٍّ بلون يماثلُه في الوحدة اللُّغوِيَّة، وهذا من الممكن أن يسهم في تعليم اللُّغة العربيَّة للناطقين بها أو بغيرها، بعد أن عجزت الوسائل التَّقليدِيَّة في التَّصدِّي لمشكلة تعليم العربيَّة، ويسعفني المقام أنْ أذكر ما خلص إليه الدُّكتور نبيل على إلى مسلَّمة مفادها «أن ارتباط تعليم العربية بالحاسوب سيمحو من أذهان كثيرٍ من صغارنا –بل كبارنا أيضا– مفهومًا خاطئًا أوشك أنْ يُرَسَّخ في أذهانهم مؤداه أن العربيَّة لغة قديمة بالية، وعاجزة عن مواكبة التَّطوُّر في جميع المجالات الاجتماعيَّة الحديثة»[١].

٢-١١ فئة سَلْمة الكلامية SALMA Tag set

قدم الدكتور مجدي صوالحة ضمن أطروحته للدكتوراة التي أعدها بجامعة ليدز فئة كلامية تنطلق من السمات الصرفية الخالصة للغة العربية، بغرض بناء معنون لأقسام الكلام، يسهم في عنونة المدونات اللغوية بالمعلومات الصرفية لكل كلمة.[٢]

وبلغ عدد فئات السمات الصرفية ٢٢ فئة، مستندًا في تغطيتها على الكتب التراثية العربية، بحيث تغطي السمات الصرفية العربية الخالصة التي لا يشوبها تأثر باللغة الإنجليزية كما ظهر في العديد من الفئات الكلامية الأخرى.[٣]

وتتم العنونة في صور متجهات متتالية، تشغل كل فئة موقعها في المتجه المحدد سابقًا، وذلك بغرض المحاذاة التي تقتضيها تقنيات الذكاء الاصطناعي. ويوضح الشكل رقم ٢٤ مثالًا معنونًا بفئة صوالحة الكلامية للآية القرآنية ”وَوَصَّيْنَا الْإِنْسَانَ بِوَالِدَيْهِ إِحْسَانًا“.

Word	Morphemes				Tag
wa waaṣṣaynā And We have enjoined	ووصّينا	وَ	*wa*	*And*	p--c------------------
		وَصّيْ	*waṣṣay*	*Have enjoined*	v-p---mpfs-s-amohvtt&-
		نا	*nā*	*We*	r---r-xpfs-s----hn----
al-'insāna (on) man	الإنسان	أل	*al-*	*The*	r--d------------------
		إنسان	*'insāna*	*man*	nq----ms-pafd---htbt-s
bi- wālidayhi His parents	بوالديه	بِ	*bi*	*To*	p--p------------------
		والِد	*wālida*	*Parents*	nu----md-vgki---htot-s
		يْ	*y*	*Both*	r---r-xdts-s----------
		هِ	*hi*	*His*	r---r-msts-k----------
ḥusn^an Kindness	حُسناً	حُسْن	*ḥusn*	*kindness*	ng----ms-vafi---ndst-s
		اً	*an*		r---k------f----------

الشكل: ٢٤ مثال معنون بفئة سلمة الكلامية[٤]

---

(١) على (نبيل): اللغة العربية والحاسوب، ص ٩.

(2) Sawalha, M. (2011). Ibid. P.123.

(3) Sawalha, M. (2011). Ibid. P.123.

(4) Sawalha, M. (2011). Ibid. P.124.

ويوضح الجدول رقم ٢١ فئات سلمة الكلامية التي تتكون من ٢٢ سمة صرفية:

Position	Morphological Features Categories		
1	Main Part-of-Speech	أقسام الكلام الرئيسيّة	*'aqsām al-kalām ar-ra'īsiyya'*
2	Part-of-Speech: Noun	أقسام الكلام الفرعيّة (الاسم)	*'aqsām al-kalām al-far'iyya' (al-'ism)*
3	Part-of-Speech: Verb	أقسام الكلام الفرعيّة (الفعل)	*'aqsām al-kalām al-far'iyya' (al-fi'l)*
4	Part-of-Speech: Particle	أقسام الكلام الفرعيّة (الحرف)	*'aqsām al-kalām al-far'iyya' (al-ḥarf)*
5	Part-of-Speech: Other (Residual)	أقسام الكلام الفرعيّة (أخرى)	*'aqsām al-kalām al-far'iyya' ('uḫrā)*
6	Punctuation marks	أقسام الكلام الفرعيّة (علامات الترقيم)	*'aqsām al-kalām al-far'iyya' ('alāmāt at-tarqīm)*
7	Gender	المذكّر والمؤنّث	*al-muḏakkar wa al-mu'annaṯ*
8	Number	العدد	*al-'adad*
9	Person	الاسناد	*al-'isnād*
10	Inflectional morphology	الصّرف	*aṣ-ṣarf*
11	Case or Mood	الحالة الإعرابية للاسم أو الفعل	*al-ḥāla^{tu} al-'i'rābiyya^{tu} lil-'ism 'aw al-fi'l*
12	Case and Mood marks	علامة الإعراب أو البناء	*'alāmāt al-'i'rāb wa al-binā'*
13	Definiteness	المعرفة والنّكرة	*al-ma'rifa^{t} wa an-nakira^{t}*
14	Voice	المبني للمعلوم و المبني للمجهول	*al-mabnī lil-ma'lūm wa al-mabnī lil-mağhūl*
15	Emphasized and non-emphasized	المؤكّد وغير المؤكّد	*al-mu'akkad wa ğayr al-mu'akkad*
16	Transitivity	اللازم والمتعدي	*al-lāzim wa al-muta'addi*
17	Rational	العاقل وغير العاقل	*al-'āqil wa ğayr al-'āqil*
18	Declension and Conjugation	التصريف	*at-taṣrīf*
19	Unaugmented and Augmented	المجرّد والمزيد	*al-muğarrad wa al-mazīd*
20	Number of root letters	عدد أحرف الجذر	*'adad 'aḥruf al-ğaḏr*
21	Verb root	بنية الفعل	*bunya^{tu} al-fi'l*
22	Noun finals	أقسام الأسم تبعاً للفظ آخره	*'aqsām al-'ismi tib$^{'an}$ li-lafẓi 'āḫirhi*

الجدول: ٢١ فئة سلمة الكلامية[1]

٣. عنونة المدونة اللغوية – عينة الدراسة – بالأقسام الكلامية

ولما كان البناء الشجري النحوي يقتضي تحليل العلاقات التركيبية في أبنية الجمل، فإن من البديهي أن ينطلق التحليل من تحديد الأجزاء الكلامية لتلك الأبنية التي تعدُّ من أهم المدخلات الابتدائية لأي عملية تحليل نحوي/ تركيبي.

وسعيًا إلى توفير أو تصميم فئة للأجزاء الكلامية يُعتمد عليها في عنونة المدونة اللغوية عينة الدراسة، تقترح الدراسة أن تسير على فئة (آر دي آي) للأنواع الكلامية مع الإضافة والحذف لبعض السمات. وقد تم الاعتماد على هذه الفئة الكلامية – دون غيرها – لأسباب منها:

(1) Sawalha, M. (2011). Ibid. P.126.

• تستوعب هذه الفئة طبيعة الكلمة التركيبية؛ لأنها تقوم على المتجهات الكلامية لأجزاء الكلمة.

• تنبثق من أصول النحو العربي، بعكس العديد من الفئات العربية المنبثقة من فئات اللغة الإنجليزية.

• التقاؤها مع ما تهدف إليه الدراسة، حيث تم توظيفها لاستنباط علامات الضبط للعربية المكتوبة غير المشكولة، وهو ما يناط به التحليل النحوي.

وانطلاقا مما تسعى إليه الدراسة، فقد اقترحت الدراسة سمتين إضافيتين للفئة الكلامية مع حذف أربع سمات منها، أما السمات التي أضيفت إلى الفئة الكلامية فهي "علامة الترقيم" ، و"العدد" ؛ لما لهما من دور تأثيري في الاستدلال على العلاقات التركيبية. بينما السمات التي حذفت من الفئة الكلامية هي سمتا بداية الكلمة SOW والحشو Padding اللتان وضعتا لتهيئة قاعدة البيانات اللغوية على التدريب أثناء بناء معنون (آر دي آي) للأجزاء الكلامية المسمَّى بـ (السِّراج) [١]، وسمتا الزوائد النحوية "همزة الاستفهام" و"ضمير النصب أو الجر" اللتان كانتا يعنيان بزوائد الكلمة اللصقية.

وقد استعانت الدراسة بمعنون السراج للأجزاء الكلامية[٢] باعتبارها أداة مساعدة؛ لعنونة نصوص المدونة اللغوية بالأجزاء الكلامية آليًّا، ثم مراجعة المخرج يدويًّا مع إضافة بعض التعديلات التي تتفق والهدف المنشود من البناء.

ومن أمثلة مخرجات الآلية التي يستند إليها البحث في عنونة الأجزاء الكلامية لهذه الفقرة المقتبسة من المدونة اللغوية: " فحصت د. زينب الشريف ثلاثين عينة عسل، تم سحبها من جميع محافظات مصر، وجميع ها ظهر بها المادة الملوثة. و أشارت الباحثة ب قسم الكيمياء الحيوية ب هيئة الرقابة و البحوث الدوائية المصرية، أن هـ على الرغم من الحساسية الضعيفة ل الأجهزة الموجودة داخل الهيئة، ف إن النتائج كل ها جاءت تؤكد وجود الكلورامفينيكول في ١٠٠٪ من عينات العسل." إسلام أون لاين– ٢٠٠٨ م

(١) المعنون النحوي للمفردات العربية، للاطلاع والتجريب، يمكن زيارة موقع (آر دي آي).

(٢) قد تعدت نسبة دقته ٩٦٪ من خلال فحص عينة من ١٠٠٠ كلمة.

{(فحصت)}; (NullPrefix) 1 (Active) 12 (Verb) 37 (Past) 39 (Femin) 48 (Single) 51} {(د.)}; (NullPrefix) 1 (Active) 12 (Verb) 37 (PresImperat) 40 (NullSuffix) 46}. {(زينب)}; (NullPrefix) 1 (Noun) 14 (NullSuffix) 46} {(الشريف)}; (Definit) 7 (Noun) 14 (ExaggAdj) 18 (NullSuffix) 46} {(ثلاثين)}; (NullPrefix) 1 (Noun) 14 (Plural) 42 (Masc) 50 (NonAdjunct) 54 (MANSS_MAGR) 55} {(عينة)}; (NullPrefix) 1 (Noun) 14 (ExaggAdj) 18 (Femin) 48 (Single) 51} {(عسل)}; (NullPrefix) 1 (Noun) 14 (NullSuffix) 46}، {(تم)}; (NullPrefix) 1 (Active) 12 (Verb) 37 (Past) 39 (NullSuffix) 46} {(سحبها)}; (NullPrefix) 1 (Noun) 14 (ObjPossPro) 57} {(من)}; (NullPrefix) 1 (Prepos) 5 (NullSuffix) 46} {(جميع)}; (NullPrefix) 1 (Noun) 14 (ExaggAdj) 18 (NullSuffix) 46} {(محافظات)}; (NullPrefix) 1 (Noun) 14 (Plural) 42 (Femin) 48} {(مصر)}; (NullPrefix) 1 (Noun) 14 (NullSuffix) 46}، {(وجميعها)}; (Conj) 3 (Noun) 14 (ExaggAdj) 18 (ObjPossPro) 57} {(ظهر)}; (NullPrefix) 1 (Active) 12 (Verb) 37 (Past) 39 (NullSuffix) 46} {(بها)}; (NullPrefix) 1 (PrepPronComp) 22 (NullSuffix) 46} {(المادة)}; (Definit) 7 (Noun) 14 (SubjNoun) 17 (Femin) 48 (Single) 51} {(الملوثة)}; (Definit) 7 (Noun) 14 (SubjNoun) 17 (Femin) 48 (Single) 51}، {(وأشارت)}; (Conj) 3 (Active) 12 (Verb) 37 (Past) 39 (Femin) 48 (Single) 51} {(الباحثة)}; (Definit) 7 (Noun) 14 (SubjNoun) 17 (Femin) 48 (Single) 51} {(بقسم)}; (Prepos) 5 (Noun) 14 (NullSuffix) 46} {(الكيمياء)}; (Definit) 7 (Noun) 14 (NoSARF) 21 (NullSuffix) 46} {(الحيوية)}; (Definit) 7 (Noun) 14 (RelAdj) 47 (Femin) 48 (Single) 51} {(بهيئة)}; (Prepos) 5 (Noun) 14 (Femin) 48 (Single) 51} {(الرقابة)}; (Definit) 7 (Noun) 14 (NounInfinit) 15 (Femin) 48 (Single) 51} {(والبحوث)}; (Conj) 3 (Definit) 7 (Noun) 14 (NullSuffix) 46} {(الدوائية)}; (Definit) 7 (Noun) 14 (NounInfinit) 15 (RelAdj) 47 (Femin) 48 (Single) 51} {(المصرية)}; (Definit) 7 (Noun) 14 (RelAdj) 47 (Femin) 48 (Single) 51}، {(أنه)}; (NullPrefix) 1 (ParticleNAASIKH) 34 (ObjPossPro) 57} {(على)}; (NullPrefix) 1 (Prepos) 5 (NullSuffix) 46} {(الرغم)}; (Definit) 7 (Noun) 14 (NullSuffix) 46} {(من)}; (NullPrefix) 1 (Prepos) 5 (NullSuffix) 46} {(الحساسية)}; (Definit) 7 (Noun) 14 (NounInfinit) 38 (Femin) 48 (Single) 51} {(الضعيفة)}; (Definit) 7 (Noun) 14 (ExaggAdj) 18 (Femin) 48 (Single) 51} {(للأجهزة)}; (Prepos) 5 (Definit) 7 (Noun) 14 (Plural) 42 (Femin) 48 (Single) 51} {(الموجودة)}; (Definit) 7 (Noun) 14 (ObjNoun) 19 (Femin) 48 (Single) 51} {(داخل)}; (NullPrefix) 1 (Noun) 14 (SubjNoun) 17 (NullSuffix) 46} {(الهيئة)}; (Definit) 7 (Noun) 14 (Femin) 48 (Single) 51}، {(فإن)}; (Conj) 3 (ParticleNAASIKH) 34 (NullSuffix) 46} {(النتائج)}; (Definit) 7 (Noun) 14 (NoSARF) 21 (Plural) 42 (NullSuffix) 46} {(كلها)}; (NullPrefix) 1 (Noun) 14 (ObjPossPro) 57} {(جاءت)}; (NullPrefix) 1 (Active) 12 (Verb) 37 (Past) 39 (Femin) 48 (Single) 51} {(تؤكد)}; (Present) 10 (Active) 12 (Verb) 37 (NullSuffix) 46} {(وجود)}; (NullPrefix) 1 (Noun) 14 (NullSuffix) 46} {(الكلورامفينيكو)}; (Translit) 61} {(في)}; (NullPrefix) 1 (Prepos) 5 (NullSuffix) 46} 100% {(من)}; (NullPrefix) 1 (Prepos) 5 (NullSuffix) 46} {(عينات)}; (NullPrefix) 1 (Noun) 14 (ExaggAdj) 18 (Plural) 42 (Femin) 48} {(العسل)}; (Definit) 7 (Noun) 14 (NullSuffix) 46}.

شكل:٢٥ مخرجات آلية السراج للعنونة الكلامية

وتحقيقًا للغاية المنشودة من الدراسة، فكان لابد من تناول بعض الإجراءات العملية – المتضمنة الحذف والزيادة في سمات الفئة الكلامية – على مخرجات هذه الآلية:

• مراجعة الأجزاء الكلامية لمفردات المدونة اللغوية – المعنونة آليًّا – وتصحيحها من الناحية اللغوية.

• إزالة الأرقام التي تجاور كل سمة من السمات الكلامية كما في المثال السابق؛ لأنها ببساطة هي مجرد تعبير عن مورفيمات أجزاء الكلام، ألزمت كل نوع كلامي للتوازي بين مورفيمات المحلل الصرفي (الميزان) ومورفيمات الأجزاء الكلامية أثناء تصميم المعنون الآلي السراج؛ ولذلك لم يكن لوجودها فائدة في بنائنا النحوي.

• ولما كانت الفئة الكلامية المستخدمة في الدراسة التي ينتهجها المعنون الكلامي تقوم على النص غير المجزئ، فكان لابد من إجراء بعض التعديلات عليها والتي تمثلت في إدماج سمتي «الهمزة الاستفهامية» و«الأداة الاستفهامية» في سمة واحدة باعتبارهما وحدة نحوية واحدة، وحذف سمة « ضمير نصب أو جر» التي كان مردُّها المورفيمات اللصقية.

• ولما كانت نصوص المدونة اللغوية المجزأة ستكون مُدخلًا[1] للمحلل النحوي Stanford Parser – المستخدم في الدراسة الذي يتضمن محللًا للأقسام الكلامية، فكانت المعالجة اليدوية لمخرجات المعنون الكلامي من حيث فصل الزوائد منها خير سبيل لحل هذه المعضلة، ومن ثم المقابلة بالتوازي Mapping بين مخرجات العنونة التي تم تعديلها ومخرجات عنونة الأجزاء الكلامية للمحلل النحوي.

وجدير بالذكر بأنه قد تم تطوير أداة برمجية – من قبل الباحث – لإجراء هذه المطابقة آليًّا مع إزالة الأقواس التي تحيط كل جزء كلامي ثم إضافة علامة الجمع "+" بين هذه الأجزاء الكلامية. ويوضح الشكل التالي شفرة «كود» الأداة المكتوبة بلغة بايثون البرمجية:

(١) المحلل النحوي يتعامل مع النص الخام فحسب.

```
#!/usr/bin/python
#coding=utf-8
import sys
import re
import codecs

def cleanAndArrangeLine(line):
	newWord= ‹ (‹
	line = line.strip()
	line = line[1:-1]
	words = line.split(‹ ‹)
	isEntered = True
	for word in words[1:-1]:
		if word is ‹ ›:
			continue
		if isEntered:
			newWord  = newWord+››+word
			isEntered = False
		else:
			newWord  = newWord+›+›+word
	newWord = newWord+› ‹+words[0]+›)›
	print newWord
	return newWord


def getTreeFromLine(line):
	tree = {};
	lines = re.findall(‹\(.*?\)›,line.strip());
	for line in lines:
		word = line.split(‹ ‹)[0][1:].strip().
replace(‹ُ›,››).replace(‹َ›,››).replace(‹ِ›,››).
replace(‹ٌ›,››).replace(‹ٍ›,››).replace(‹ً›,››).
replace(‹ّ›,››).replace(‹ْ›,››).replace(‹ا‹,‹أ›).
replace(‹ا‹,‹إ›).replace(‹ا‹,‹آ›).strip()
		line = cleanAndArrangeLine(line)
		tree[word] = line.strip()
	return tree

treeFilePath = sys.argv[1]
textFilePath = sys.argv[2]
outputFilePath = sys.argv[3]

treeFile = open(treeFilePath,›r›)
textFile = open(textFilePath,›r›)
textFileLines = textFile.readlines()
treeFileLines = treeFile.readlines()
outputFile = codecs.open(outputFilePath,›w›,encoding=›u
tf-8›)
```

```
count = 0
for line in textFileLines:
      line = line.replace(‹ (‹ , ››).replace(‹) ›, ››).
replace(‹ ; ›, ››).replace(‹ \» ›, ››)
      line = line.replace(‹ {‹ , ›(‹).replace(‹ } ›, ›) ›)
      tree = getTreeFromLine(line)
      lineOfTree = unicode(treeFileLines[count], ›utf-8 ›)
      lineOfTree = lineOfTree[::-1]
      for key in tree:
            key1 = unicode(key, ›utf-8 ›)
            key1 = key1[::-1]
            replacement = unicode(tree[key], ›utf-8 ›)[::-
1];
            lineOfTree = re.sub(‹ \) ›+key1+ ›.*?\
(‹ ,replacement,lineOfTree)
      count = count + 1;
      lineOfTree = lineOfTree[::-1]
      outputFile.write(lineOfTree+ ›\n ›);

treeFile.close()
textFile.close()
outputFile.close()
```

وبتطبيق الإجراءات العملية السابقة على مخرجات العنونة الآلية، يظهر المثال السابق بهذا الشكل رقم ٢٦:

(NullPrefix+Active+Verb+Past+Femin+Single فحصتْ) (NullPrefix+Noun+NullSuffix د.)
(NullPrefix+Noun+NullSuffix زَيْنبُ) (Definit+Noun+ExaggAdj+NullSuffix الشّرِيفِ)
(NullPrefix+Noun+Plural+Masc+NonAdjunct+MANSS_MAGR ثلاثِين)
(NullPrefix+Noun+ExaggAdj+Femin+Single عيِّنة) (NullPrefix+Noun+NullSuffix عسلٍ) (PUNC
,) (NullPrefix+Active+Verb+Past+NullSuffix تمّ) (NullPrefix+Noun+NullSuffix سحْبُ)
(NullPrefix+Noun+Pronoun+NullSuffix ها) (NullPrefix+Prepos+NullSuffix مِنْ)
(NullPrefix+Noun+ExaggAdj+NullSuffix جمِيعِ) (NullPrefix+Noun+Plural+Femin مُحافظاتِ)
(NullPrefix+Noun+NullSuffix مِصْر) (PUNC ،) (NullPrefix+Conj+NullSuffix و)
(NullPrefix+Noun+ExaggAdj+NullSuffix جمِيعِ) (NullPrefix+Noun+Pronoun+NullSuffix ها)
(NullPrefix+Active+Verb+Past+NullSuffix ظهر) (NullPrefix+PrpPronComp+NullSuffix بِها)
(Definit+Noun+SubjNoun+Femin+Single الْمادّةُ) (Definit+Noun+SubjNoun+Femin+Single
الْمُلوِّثةُ) (PUNC .)
(NullPrefix+Conj+NullSuffix و) (NullPrefix+Active+Verb+Past+Femin+Single أشارتْ)
(Definit+Noun+SubjNoun+Femin+Single الْباحِثةُ) (NullPrefix+Prepos+NullSuffix ب)
(NullPrefix+Noun+NullSuffix قِسْمِ) (Definit+Noun+NoSARF+NullSuffix الْكِيمْياءِ)
(Definit+Noun+RelAdj+Femin+Single الْحيويّةِ) (NullPrefix+Prepos+NullSuffix ب)
(NullPrefix+Noun+Femin+Single هيْئةِ) (Definit+Noun+NounInfinit+Femin+Single الرّقابةِ)
(NullPrefix+Conj+NullSuffix و) (Definit+Noun+NullSuffix الْبُحُوثِ)
(Definit+Noun+NounInfinit+RelAdj+Femin+Single الدّوائيّةِ)
(Definit+Noun+RelAdj+Femin+Single الْمِصْريّةِ) (PUNC ,)
(NullPrefix+ParticleNAASIKH+NullSuffix إِنّ) (NullPrefix+Noun+Pronoun+NullSuffix ه)
(NullPrefix+Prepos+NullSuffix على) (Definit+Noun+NullSuffix الرّغْمِ)
(NullPrefix+Prepos+NullSuffix مِنْ) (Definit+Noun+NounInfinit+Femin+Single الْحساسِيةِ)
(Definit+Noun+ExaggAdj+Femin+Single الضّعِيفةِ) (NullPrefix+Prepos+NullSuffix ل)
(Definit+Noun+Plural+Femin+Single الْأجْهِزةِ) (Definit+Noun+ObjNoun+Femin+Single
الْموْجُودةِ) (NullPrefix+Noun+SubjNoun+NullSuffix داخِل) (Definit+Noun+Femin+Single الْهيْئةِ)
(PUNC ,) (NullPrefix+Conj+NullSuffix ف) (NullPrefix+ParticleNAASIKH+NullSuffix إِنّ)
(Definit+Noun+NoSARF+Plural+NullSuffix النّتائِجِ) (NullPrefix+Noun+NullSuffix كُلّ)
(NullPrefix+Noun+Pronoun+NullSuffix ها) (NullPrefix+Active+Verb+Past+Femin+Single
جاءتْ) (Present+Active+Verb+NullSuffix تُؤكِّدُ) (NullPrefix+Noun+NullSuffix وُجُود) (Translit
الْكُلُورامْفِينِيْكُو) (NullPrefix+Prepos+NullSuffix فِي) (NullPrefix+CardNumber+NullSuffix 100%)
(NullPrefix+Prepos+NullSuffix_CON مِنْ) (NullPrefix+Noun+ExaggAdj+Plural+Femin عيِّناتِ)
(Definit+Noun+NullSuffix الْعسلِ) (PUNC__ .)

شكل:٢٦ مخرجات آلية السراج للعنونة الكلامية بعد تطبيق الإجراءات العملية

وبملاحظة المثال السابق يتبين مواضع الاختلاف بين العنونة الآلية والإجراءات التي تمت عليها، وقد أوضحتها في صورة بارزة في الشكل السابق:

- تصحيح عنونة اختصار لقب "دكتور"، حيث يعْتبِر المعنون الآلي كل حرف منفصل فعل أمر أو حرف غير عامل؛ لأنه لا يعتمد على النص المجزأ بشكل عام.
- عنونة علامات الترقيم والأعداد اللتين يتجاهلهما المعنون الآلي.
- عنونة الزوائد النحوية منفردة، حيث كانت متصلة بالمفردات في العنونة الآلية.
- عنونة الضمائر المجزأة التي تكون في حكم الاتصال بـعلامة Pronoun.

٣-١ أقسام الكلام Part Of Speech (POS)

بعد الإجراءات العملية التي تمت على الفئة الكلامية المستخدمة في إطار تحقيق الهدف المنشود من الدراسة، تمثلت الفئة الكلامية بعد إعادة هيكلتها - من ناحية توزيع الأقسام الكلامية الرئيسية وعدد سماتها - في ستة أقسام رئيسية (الاسم - الفعل - الأداة - الظرف - الكلمة الأجنبية - علامة الترقيم) تتخللها ٥٦ سمة.

الأقسام الرئيسية	المصطلح العربي		المصطلح الإنجليزي	الاختصار
الاسم	الاسم			
	اسم		Nominal	Noun
	مصدر		Nouns made of infinitives	NounInfinit
	اسم مصدر		NounInfinit" like"	NounInfinitLike
	اسم فاعل		Subject noun	SubjNoun
	صيغة مبالغة		Exaggeration adjective	ExaggAdj
	اسم مفعول		Object noun	ObjNoun
	اسم زمان أو مكان		Noun of time or location	TimeLocNoun
	الضمير	الضمير	pronoun	Pronoun
		ضمير رفع	Subject form pronoun	SubjPro
	اسم موصول		Relative pronoun	RelPro
	اسم إشارة		Demonstrative pronoun	DemoPro
	العدد		Cardinal number	CD

الأقسام الرئيسية	المصطلح العربي		المصطلح الإنجليزي	الاختصار
الاسم	خصائص الاسم الصرف-نحوية			
	علامات اللواصق	لا سابق	Null prefix	NullPrefix
		لا لاحق	Null suffix	NullSuffix
	علامات الصرف	ممنوع من الصرف	An Arabic feature of a specific class of nouns	NoSARF
	علامات التعريف	"ال" التعريف	Definitive article	Definit
	علامات الجنس	تأنيث	Feminine	Femin
		مذكر	Masculine	Masc
	علامات العدد	مفرد	Singular	Single
		مثنى	Binary	Binary
		جمع	Plural	Plural
	علامات للدلالة	نسب	Relative adjectives maker	RelAdj
		مضاف	Adjunct	Adjunct
		غير مضاف	NonAdjunct	NonAdjunct
	علامات الإعراب	مرفوع	1st Arabic syntactic case	MARF

الأقسام الرئيسية	المصطلح العربي		المصطلح الإنجليزي	الاختصار
الاسم	علامات الإعراب	منصوب	2nd Arabic syntactic case	MANSS
		منصوب أو مجرور	2nd or 3rd Arabic syntactic case	MANSS_MAGR
		مجرور	3rd Arabic syntactic case	MAGR
الفعل			فعل	
	فعل		Verb	Verb
	الفعل التام	الفعل اللازم	Transitive verb	Transitive
		الفعل المتعدي	InTransitive verb	InTransitive
	مبني للمعلوم (للفاعل)		Active sound	Active
	مبني للمجهول (للمفعول)		Passive sound	Passive
	الفعل الناقص	الفعل الناسخ	A class of auxiliary verbs that make the predicate of the conse-quent verbal sentence in 2nd Arabic syntac-tic case	VerbNAASIKH

الأقسام الرئيسية	المصطلح العربي		المصطلح الإنجليزي	الاختصار
الفعل	خصائص الفعل الصرف – نحوية			
	علامات اللواصق	لا سابق	Null prefix	NullPrefix
		لا لاحق	Null suffix	NullSuffix
		مضارع	Present tense	Present
		استقبال	Future tense	Future
	الزمن	ماض	Past tense	Past
		مضارع أو أمر	Present tense, or imperative	PresImperat
		أمر	Imperative	Imperative
	الجنس	تأنيث	Feminine	Femin
		مذكر	Masculine	Masc
	العدد	مفرد	Singular	Single
		مثنى	Binary	Binary
		جمع	Plural	Plural
	الإعراب	مرفوع	1st Arabic syntactic case	MARF
		منصوب	2nd Arabic syntactic case	MANSS
		منصوب أو مجزوم	2nd or 4th Arabic syntactic case	MANS_MAJZ
		مجزوم	4th Arabic syntactic case	MAJZ

الأقسام الرئيسية	المصطلح العربي	المصطلح الإنجليزي	الاختصار
الأداة	حرف جر	Preposition	Prepos
	حرف نداء	Interjection	Interj
	الجار والمجرور	Preposition-Pronoun Compound	PrepPronComp
	حرف ناسخ	A class of particles that make the subject of the consequent nominal sentence in 2nd Arabic syntactic case	ParticleNAASIKH
	العطف	Conjunctive	Conj
	لام التوكيد	Confirmation by Laam	Confirm
	أداة استفهام	InterrogArticle	InterrogArticle
	جازمة	For specific articles that make the conse-quent verb in the 4th Arabic syntactic case	JAAZIMA
	شرطية جازمة	Feature of a class of Arabic conditionals	CondJAAZIMA
	شرطية غير جازمة	Feature of a class of Arabic conditionals	CondNot JAAZIMA
	استثناء	Article of exception	Except
	لات	Arabic specific article	LAATA

الأقسام الرئيسية		المصطلح العربي	المصطلح الإنجليزي	الاختصار
الأداة		لا	Arabic specific article	LAA
		غير عاملة	A class of articles that have no syntactic effect	NoSyntaEffect
		ناصب	Arabic specific class of particles that make the consequent verb in 2nd Arabic syntactic case	ParticleNAASSIB
		مصدرية	Arabic specific article	MASSDARIYYA
	حصائص الأداة الصرفية			
	علامات اللواصق	لا سابق	Null prefix	NullPrefix
		لا لاحق	Null suffix	NullSuffix
الظرف		ظرف	Feature for certain kind of Arabic adverbs	DZARF
	حصائص الأداة الصرفية			
	علامات اللواصق	لا سابق	Null prefix	NullPrefix
		لا لاحق	Null suffix	NullSuffix
الكلمة الأجنبية		كلمة غير عربية	Translit	Translit
الترقيم		الترقيم	Punctution	Punc

الجدول: ٢٢ الفئة الكلامية بعد إعادة هيكلتها

سأتناول عناصر هذه الفئة الكلامية بشيء من التفصيل، مبتدأً بتعريف الأقسام الكلامية الرئيسية (الاسم، الفعل، الأداة، الظرف، الكلمة الأجنبية، علامة الترقيم) ثم توضيح السمات التي تندرج تحت كل قسم كلامي، مبينًا أوجه علاقتها بذلك القسم.

٣-١-١ الاسم

هو ما دل على معنى في نفسه غير مقترن بزمان: إنسان، رجل. ويتميز الاسم في القواعد العربية بموضعين مختلفين: أنواعه وخصائصه.[١]

٣-١-١-١ أنواعه:

٣-١-١-١-١ متصرف، يثنّى ويجمع ويصغر وينسب إليه.

٣-١-١-١-٢ غير متصرف، يلزم حالة واحدة.

٣-١-١-١-١متصرف، وهو نوعان:

٣-١-١-١-١-١ جامد: اسم متصرف غير مأخوذ من غيره أي أنه وضع على صورته الحالية ابتداءً؛ فليس له أصل يرجع إليه أو يتفرع منه. ويشمل اسم الذات أي اسمي الجنس (امرأة، رجل، عرب، أم) والعلم (مصر، أحمد، مريم)، واسم المعنى أي المصدر المجرد (شُرْبٌ)،[٢] وقد تم توصيفهم برمز الاسم Noun.

٣-١-١-١-١-٢ مشتق: اسم متصرف مأخوذ من غيره، له أصل يرجع إليه ويتفرع منه.[٣] ويشمل الاسم المشتق:

- **مصدر NounInfinint**

اسم يدل على حدث خاليًا من الدلالة على الزمن، متضمنًا أحرف فعله لفظًا أو تقديرًا: مثل اجتماع من الفعل اجتمع.[٤]

- **اسم مصدر NounInfinitLike**

لفظ يدل على معنى المصدر وينقص عن حروف فعله لفظا وتقديرا دون تعويض

---

(١) الدحداح (إنطوان): في لغة النحو العربي، ص ١٥.

(٢) المصدر السابق، ص ١٨.

(٣) المصدر السابق، ص ٢٧.

(٤) المصدر السابق، ص٤٤

هذا من ناحية اللفظ، أما من ناحية المعنى فهو ما تعلق بالأثر المترتب على الفعل. مثال اسم المصدر عطاء للفعل أعطى[١].

- **اسم فاعل SubjNoun**

اسم مشتق يدل على الحدث وفاعله. مثال اسم الفاعل طاعن للفعل طعن.[٢]

- **صيغة مبالغة ExaggAdj**

اسم مشتق بمعنى اسم الفاعل يدل على زيادة الوصف في الموصوف. مثال صيغة المبالغة حذر، خبير. [٣]

- **اسم مفعول ObjNoun**

اسم مشتق يدل على ما وقع عليه الفعل. مثال اسم المفعول مسؤول، مغلق.[٤]

- **اسم زمان أو مكان TimeLocNoun**

اسم مشتق يدل على وقت وقوع الفعل أو موضعه. مثال اسم الزمان أو المكان مغرب، مرمى.

٣-١-١-١-٢ غير متصرف، ويشمل:

- **الضمير**
- **الضمير pronoun**

هو الضمير الذي يكون في حكم الاتصال بآخر الكلمة قبل تجزئة الوحدات الفرعية مثل: ضمير الجر أو النصب.

- **ضمير رفع SubjPro**

هو ضمير بارز، محله من الإعراب الرفع مطلقًا. مثال ذلك الضمير المنفصل: هو، أنت، أنا، والضمير المتصل: التاء، النون، الواو، الألف. [٥]

- **اسم موصول RelPro**

اسم يدل على معيّن، ولا يتم معناه إلا بجملة تذكر بعده تسمى صلة الموصول أو الجملة المكملة للجملة الكبرى. مثال ذلك الموصول الخاص الذي له صيغ للمذكر

(١) المصدر السابق، ص ٢٨

(٢) حسن (عباس): النحو الوافي، دار المعارف ، الجزء ٣، ص ٢٣٨.

(٣) ينظر: المصدر السابق، ص ٢٥٧.

(٤) المصدر السابق، ص ٢٧١.

(٥) الدحداح (إنطوان): في لغة النحو العربي، ص ١٨٦ .

والمؤنث والمفرد والمثنى والجمع: الذي، التي،... والمشترك الذي يبقى بلفظ واحد مع الجميع: منْ، أيُّ.... [١]

• **اسم إشارة DemoPro**

اسم يدل على معيّن بالإشارة إليه سواء أكان إنسانًا أو حيوانا أو شيئًا أو مكانًا. مثال ذلك إشارة القريب "هذا، هذه، هؤلاء"، وإشارة البعيد "ذلك، تلك، أولئك". [٢]

• **العدد CD**

اسم يدل على العدد الترتيبي الرقمي، مثل: الأرقام العددية (٣, ١٢)، التواريخ (٥-٦-٢٠٠٩).

٣-١-١-٢ خصائص الاسم الصرف-نحوية:

تندرج خصائص الاسم في عدد من العلامات:

٣-١-١-٢-١ علامات اللواصق

• **لا سابق NullPrefix**

علامة تستخدم للدلالة على عدم وجود سابق، يسبق جذع الكلمة (الاسم).

• **لا لاحق NullSuffix**

علامة تستخدم للدلالة على عدم وجود لاحق، يلحق جذع الكلمة (الاسم).

٣-١-١-٢-٢ علامات الصرف

• **ممنوع من الصرف NoSARF**

هو الاسم المعرب الذي لايدخله التنوين في آخره، وتكون الفتحة علامةً لجره. مثال ذلك مررت بسليمان. [٣]

٣-١-١-٢-٣ علامات التعريف

• **أداة التعريف Definit**

حرف تعريف يدخل على الاسم النكرة ليعرفه. مثال ذلك "أل" في كلمة "الرجل".

٣-١-١-٢-٤ علامات الجنس

• **مذكر Masc**

(١) المصدر السابق، ص ٣٣.

(٢) المصدر السابق، ص ١٦.

(٣) المصدر السابق، ص ٣١٨.

هو ما يمكن الإشارة إليه باسم الإشارة المذكر سواء أكان مفردًا أو مثنى أو جمعًا. مثال ذلك"عيسى".

- **مؤنث Femin**

هو ما يمكن الإشارة إليه باسم الإشارة المؤنث سواء أكان مفردًا أو مثنى أو جمعًا. مثال ذلك"أمينة".

٣-١-١-٢- ٥ علامات العدد

- **مفرد Single**

وضع صرفي للاسم الذي يدل على واحد.

- **مثنى Binary**

وضع صرفي يدل على اثنين من الاسم المفرد.

- **جمع Plural**

وضع صرفي يدل على أكثر من اثنين.

٣-١-١-٢-٦ علامات للدلالة على ما يلي:

- **نسب RelAdj**

إلحاق ياء مشددة بالاسم للدلالة على انتساب الشيء إليه.

- **مضاف Adjunct**

هي أسماء ناقصة الدلالة بنفسها؛ لذا لا تنفك عن الإضافة أي تحتاج إلى مضاف إليه يوضح معناها. مثال ذلك أي، إزاء، أمام، تجاه، تحت، قبل، فوق....

- **غير مضاف NonAdjunct**

هي الأسماء التي تحتوي في آخرها على نون التثنية وجمع المذكر السالم، وملحقاتها. مثال ذلك (الجنسيْنِ، المحترفين، اثنان، أهلون).

٣-١-١-٢-٧ علامات الإعراب(١)

- **مرفوع MARF**

ويقصد بها هنا بالمرفوع الأسماء التي تحمل علامات الرفع الفرعية مثل الألف في حالة المثنى، والواو في الأسماء الخمسة، والجمع المذكر السالم.

---

(١) لما كان التعامل مع النص غير المشكول فحسب، فكان الاقتصار على العلامات البارزة (الحروف) في هيئة الكلمة المتمثلة في إعراب المثنى وجمع المذكر السالم، وملحقاتهما.

• منصوب MANSS

ويقصد بها هنا بالمنصوب الأسماء التي تحمل العلامات القطعية الدالة على النصب مثل الألف في الأسماء الخمسة: أكرم الناس أباك.

• منصوب أو مجرور MANSS_MAGR

ويقصد بها هنا بالمنصوب أو المجرور الأسماء التي تحتمل حالتي النصب والجر، لتشابها في حمل علامة الياء والنون في المثنى وجمع المذكر السالم. مثل زرت العائلتيْنِ بالأمس، ورأيت المحترفِين في الشارع.

• مجرور MAGR

والمقصود هنا بالمجرور الأسماء التي تحمل العلامات القطعية الدالة على الجر مثل الياء في الأسماء الخمسة: مررت بأبيك.

ومن أمثلة عنونة القسم الكلامي "الاسم" في المدونة اللغوية عينة الدراسة:

الكلمة	العنونة باللغة العربية	العنونة باللغة الإنجليزية
وقت	لا سابق+اسم+لا لاحق	NullPrefix+Noun+NullSuffix
الشعار	تعريف+اسم+لا لاحق	Definit+Noun+NullSuffix
الحالة	تعريف+اسم+مؤنث+مفرد	Definit+Noun+Femin+Single
الأطباء	تعريف+اسم+ممنوع من الصرف+جمع+مذكر+لا لاحق	Definit+Noun+NoSARF+Plural+Masc+NullSuffix
العامة	تعريف+اسم+اسم فاعل+مؤنث+مفرد	Definit+Noun+SubjNoun+Femin+Single
سياسات	لاسابق+اسم+مؤنث+جمع	NullPrefix+Noun+Femin+Plural
اعتبار	لا سابق+اسم+مصدر+لا لاحق	NullPrefix+Noun+NounInfinit+NullSuffix
صناديق	لا سابق+اسم+ممنوع من الصرف+جمع+لا لاحق	NullPrefix+Noun+NoSARF+Plural+NullSuffix

الكلمة	العنونة باللغة العربية	العنونة باللغة الإنجليزية
المستثمرين	تعريف+اسم+اسم فاعل+جمع+مذكر+غير مضاف+منصوب أو مجرور	Definit+Noun+SubjNoun+Plural+Masc+NonAdjunct+MANSS+MAGR
الأقل	تعريف+اسم+ممنوع من الصرف+لا لاحق	Definit+Noun+NoSARF+NullSuffix
الذي	تعريف+اسم+اسم موصول+مذكر+مفرد+لا لاحق	Definit+Noun+RelPro+Masc+Single+NullSuffix
المصرية	تعريف+اسم+نسب+مؤنث+مفرد	Definit+Noun+RelAdj+Femin+Single
هذه	لا سابق+اسم+اسم إشارة+مذكر+مفرد	NullPrefix+Noun+DemoPro+Masc+Single
المدرسة	تعريف+اسم+اسم زمكان+مؤنث+مفرد	Definit+Noun+TimeLocNoun+Femin+Single
الضرورة	تعريف+اسم+اسم مصدر+مؤنث+مفرد	Definit+Noun+NounInfinitLike+Femin+Single
المؤتمر	تعريف+اسم+اسم مفعول+لا لاحق	Definit+Noun+ObjNoun+NullSuffix
الأخيرين	تعريف+اسم+صيغة مبالغة+مثنى+غير مضاف+منصوب أو مجرور	Definit+Noun+ExaggAdj+Binary+NonAdjunct+MANSS+MAGR
الشعاعان	تعريف+اسم+مرفوع+مثنى+غير مضاف	Definit+Noun+MARF+Binary+NonAdjunct
ذا	لا سابق+اسم+مذكر+مفرد+مضاف+منصوب+لا لاحق	NullPrefix+Noun+Masc+Single+Adjunct+ MANSS+NullSuffix
ذي	لا سابق+اسم+مذكر+مفرد+مضاف+مجرور+لا لاحق	NullPrefix+Noun+Masc+Single+Adjunct+MAGR+NullSuffix
١٩٥٩	عدد	CD

الجدول:٢٣ عنونة القسم الكلامي " الاسم " في المدونة اللغوية عينة الدراسة

٣-١-٢ الفعل

هو ما يدل على شيئين معًا، حالة أو حدث، وعلى زمن يقترن بهما: ضرب، شرب، فعل. ويتميز الفعل في القواعد بموضوعين مختلفين: أنواعه وخصائصه: [١]

٣-١-١-١ أنواعه: ينقسم الفعل إلى قسمين رئيسيين:

٣-١-١-١-١ الفعل التام[٢]

فعل يرتبط بفاعله عن طريق النسبة الإسنادية، ويتمم بذلك المعنى المقصود. وهو نوعان:

- **الفعل اللازم Transitive**

هو الفعل الذي يستقر حدوثه في فاعله، ويكتفي برفعه؛ ليتم معناه. مثال ذلك مات، حسُن، طهر.

- **الفعل المتعدي InTransitive**

هو الفعل الذي يتجاوز فاعله، وينصب مفعولًا به؛ ليستكمل معناه. مثال ذلك كتب، منح، علم.

ويُقسّم الفعل المتعدي إلى قسمين:

- **مبنيٌّ للمعلوم (للفاعل) Active**

هو الفعل المعلوم الذي يذكر معه فاعله. مثال ذلك الفعل شرِب في جملة "شرِب أحمد اللبن"

- **مبنيٌّ للمجهول (للمفعول) Passive**

هو الفعل المجهول الذي حذف فاعله، وناب عنه المفعول به. مثال ذلك الفعل شُرِب في جملة " شُرِب اللبن".

٣-١-١-١-٢ الفعل الناقص

- **فعل ناسخ VerbNAASIKH**

هو الفعل الذي يدخل على الجملة الاسمية، فيرفع المبتدأ ويسمى اسمها، وينصب الخبر ويسمى خبرها، فيُغيرهما لفظًا بنقل الإعراب من حال إلى حال، ويغيرهما معنىً

(١) المصدر السابق، ص ٢٢٨.

(٢) حسن (عباس): النحو الوافي، ص ١٥٠.

بنقل حدوث الزمن من زمنٍ لآخر، مثال ذلك" كان الطالب مجتهدًا في دراسته". [1]

٣-١-١-٢ خصائص الفعل الصرف-نحوية:

٣-١-١-٢-١ علامات اللواصق:

- **لاسابق NullPrefix**

علامة تستخدم للدلالة على عدم وجود سابق يسبق جذع الكلمة (الفعل).

- **لا لاحق NullSuffix**

علامة تستخدم للدلالة على عدم وجود لاحق يلحق جذع الكلمة (الفعل).

- **مضارع Present**

علامة تستخدم للدلالة على حرف المضارعة (الياء – التاء – الهمزة - النون) الذي يسبق جذع الفعل.

- **استقبال Future**

علامة تستخدم للدلالة على حرف الاستقبال(السين) الذي يسبق الفعل المضارع.

٣-١-١-٢-٢ الزمن:

- **ماض Past**

هو ما يدل على صيغة جذع الفعل الماضي الذي يدل على حدث حصل قبل عملية التكلم. وتتمثل صيغته في الفعل الماضي: مثل وضح، شرب، أكل.

- **مضارع أو أمر PresImperat**

هو ما يدل على صيغة جذع الفعل الأمر الذي يصاغ من الفعل المضارع المعلوم، ثم يجزم ويحذف منه حرف المضارعة مثل: دع، قل.

- **أمر Imperative**

هو ما يدل على صيغة جذع فعل الأمر الذي يصاغ من الفعل المضارع المبني المعلوم أو المعلوم فاعله إذا كان ما بعد حرف المضارع ساكنا فتزاد همزة في أوله، مثل: اُنظر، اِعلم، أكْرِم.

(١) الدحداح (إنطوان): في لغة النحو العربي، ص٢٣٩.

٣-١-١-٢-٣ الجنس

- **مذكر Masc**

هو ما يمكن تحديده من خلال مورفيمات صيغة الفعل الدالة على التذكير مثل لواحق أو سوابق الغائب المذكر أو المخاطب المذكر أو المتكلم المذكر المتصلة بالفعل. مثال ذلك الضمير في كتبت.

- **مؤنث Femin**

هو ما يمكن تحديده من خلال مورفيمات صيغة الفعل الدالة على التأنيث مثل لواحق أو سوابق الغائب المؤنث أو المخاطب المؤنث أو المتكلم المؤنث المتصلة بالفعل. مثال ذلك الضمير في كتبتِ.

٣-١-١-٢-٤ العدد:

- **مفرد Single**

هو ما يمكن تحديده من خلال صيغة الفعل المجردة من السوابق واللواحق الدالتين على التأنيث والتذكير.

- **مثنى Binary**

هو ما يمكن تحديده من خلال لواصق التثنية (الألف والنون، والياء والنون) المتصلة بالفعل.

- **جمع Plural**

هو ما يمكن تحديده من خلال لواصق الجمع المتصلة بالفعل.

٣-١-١-٢-٥ الحالة الإعرابية

- **مرفوع MARF**

ويقصد هنا بالمرفوع الأفعال التي تحمل علامات الرفع الفرعية مثل الرفع بثبوت النون نيابة عن الضمة في الأفعال الخمسة: يفعلان، يفعلون، تفعلان، تفعلون، تفعلين.

- **منصوب MANSS**

ويقصد هنا بالمنصوب الأفعال التي تحمل العلامات القطعية الدالة على النصب مثل حذف النون من الأفعال الخمسة المتصلة باللام الناصبة.

- **منصوب أو مجزوم MANS_MAJZ**

ويقصد هنا بالمنصوب أو المجزوم الأفعال المضارعة التي تتشابه في نيابتها عن

علامتها الأصلية في حالتي النصب والجزم بعلامة حذف النون في الأفعال الخمسة.

- **مجزوم MAJZ**

ويقصد هنا بالمجزوم الأفعال المضارعة التي تختص بصيغة الأمر للمخاطب. مثال ذلك اسمعْ، طالبْ.

ومن أمثلة عنونة القسم الكلامي " الفعل" في المدونة اللغوية عينة الدراسة:

الكلمة	العنونة باللغة العربية	العنونة باللغة الإنجليزية
أشار	لا سابق+ مبني للمعلوم+ فعل+ماض+لا لاحق	NullPrefix+Active+Verb+Past+NullSuffix
دعت	لاسابق+مبني للمعلوم+فعل+ماض+مؤنث+مفرد	NullPrefix+Active+Verb+Past+Femin+Single
لأوجع	توكيد+مبني للمعلوم+فعل+ماض+لا لاحق	Confirm+Active+Verb+Past+NullSuffix
ستصطدمون	استقبال+مضارع+مبني للمعلوم+فعل+مرفوع+ضمير رفع	Future+Present+Active+Verb+MARF+SubjPro
سأبدأ	استقبال+مضارع+مبني للمعلوم+فعل+لا لاحق	Future+Present+Active+Verb+NullSuffix
اعلم	لا سابق+أمر+مبني للمعلوم+فعل+لا لاحق	NullPrefix+Imperative+Active+Verb+NullSuffix
أطلقوا	لا سابق+مبني للمعلوم+فعل+ماض+ضمير رفع	NullPrefix+Active+Verb+Past+SubjPro
انقلي	لاسابق+أمر+مبني للمعلوم+فعل+مجزوم+ضمير رفع	NullPrefix+Imperative+Active+Verb+MAJZ+SubjPro
أقيمت	لاسابق+مبني للمجهول+فعل+ماض+مؤنث+مفرد	NullPrefix+Passive+Verb+Past+Femin+Single
أجري	لاسابق+مبني للمجهول+فعل+ماض+لا لاحق	NullPrefix+Passive+Verb+Past+NullSuffix

الكلمة	العنونة باللغة العربية	العنونة باللغة الإنجليزية
انضمت	لا سابق+فعل+لازم+ماض+مؤنث+مفرد	NullPrefix+Verb+Intransitive +Past+Femin+Single
أدى	لا سابق+فعل+لا لاحق	NullPrefix+Verb+NullSuffix
ليس	لا سابق+ فعل ناسخ+لا لاحق	NullPrefix+VerbNAASIKH+ NullSuffix
ليست	لا سابق+ فعل ناسخ+ماض+مؤنث+مفرد	NullPrefix+VerbNAASIKH+ Past+Femin+Single
لتنضم	ناصب+مضارع+مبني للمعلوم+فعل+لازم+لا لاحق	ParticleNAASSIB+Present+ Active+Verb+Intransitive+N ullSuffix
يدرسوا	مضارع+مبني للمعلوم+فعل+منصوب أو مجزوم+ضمير رفع	Present+Active+Verb+MAN S+MAJZ+SubjPro
يتحملون	مضارع+مبني للمعلوم+فعل+مرفوع+ضمير رفع	Present+Active+Verb+MAR F+SubjPro
انتخبوا	لا سابق+أمر+فعل+مجزوم+ضمير رفع	NullPrefix +Imperative+Verb+ MAJZ+SubjPro
قل	لا سابق+مبني للمعلوم+فعل+مضارع أو أمر+لا لاحق	NullPrefix+Active+Verb+Pre sImperat+NullSuffix

الجدول: ٢٤ عنونة القسم الكلامي "الفعل" في المدونة اللغوية عينة الدراسة

٣-١-٣ الأداة

الأداة مصطلح يؤدي معنى التعليق أو الربط، والعلاقة التي تعبر عنها الأداة إنما تكون بالضرورة بين الأجزاء المختلفة من الجملة.[١] وقد اشتملت الفئة الكلامية على:

(١) حسان (تمام): اللغة العربية معناها ومبناها، ص ١٢٣.

- **حرف جر Prepos**

هو الحرف الذي يدل على معنى في غيره، أي الاسم الذي يليه مباشرة، وذلك بجره إلى متعلقه. وحروف الجر تسعة عشر حرفا منها: في، على، بـ، لـ،...[١]

- **حرف نداء Interj**

"هو حرف معنى يستخدم لتوجيه دعوة ما أو تنبيه المخاطب". وقد اشتملت اللغة العربية على ثمانية حروف نداء منها: يا، أيا، وا، آيْ، أيْ. [٢]

- **الجار والمجرور PrePronComp**

يقصد بالجار والمجرور هنا المركب غير الكلامي الذي يتكون من حرف الجار والضمير الذي يليه باعتبارهما وحدة نحوية واحدة. مثال ذلك بها، فيها، فيه، به.

- **حرف ناسخ ParticleNAASIKH**

هو حرف معنى يدخل على الجملة الاسمية، فيحدث ضربًا من التحويلات في حركة إعراب ركنيها. وتعبّر هذه العلامة عن حروف إن وأخواتها فحسب.

- **العطف Conj**

"هو حرف معنى يستخدم لمشاركة المعطوف والمعطوف عليه في اللفظ والمعنى"، وتعبر هذه العلامة عن الحروف التالية: الواو، أو، ثم، الفاء، بل. [٣]

- **لام التوكيد Confirm**

هي لام مفتوحة غير عاملة تفيد توكيد مضمون الجملة. وتدخل على الاسم لتأكيده مثل: لمحمدٌ عظيم، وعلى ضمير الفصل مثل: محمد لهو الصادق، وعلى إن الشرطية: لئن تذاكر، ستنجح، وعلى جواب القسم لتفيد التوكيد: لقد أفصحت وقائع عام كامل.

- **أداة استفهام InterrogArticle**

هي أداة يستعلم بها عن شيء أو أمر، وتتضمن حروف الاستفهام: الهمزة، هل،

(١) الدحداح (إنطوان): في لغة النحو العربي، ص١٣٢.

(٢) المصدر السابق، ص ١٣٨.

(٣) المصدر السابق، ص ١٣٤.

وأسماء الاستفهام: كيف، متى، من، ماذا، ما ، أي، كم، أين، أيان. [١]

- **جازمة JAAZIMA**

هي الأداة التي تجزم الفعل المضارع مثل: لم، لمّا، لام الأمر، لا الناهية.

- **شرطية جازمة condJAAZIMA**

هي الأداة التي تجزم فعلين مضارعين (فعل الشرط وجوابه)، مثل: إنْ، إذما، منْ، مهما، أيّ، أينما، أيّان، أني، حيثما.

- **شرطية غير جازمة CondNotJAAZIMA**

هي أدوات تتضمن معنى الشرط ولا تجزم، وهي: إذا، لو، لولا، أمّا، لمّا، كلّما.

- **استثناء Except**

هي أداة تخرج الاسم الواقع بعدها من حكم ما قبلها، وهي: إلا، سوى، غير، بيد، عدا، لا سيما.

- **لات (نافية مشبهة بليس) LAATA**

حرف نفي مشبه بليس، يدخل على الجملة الاسمية فيرفع المبتدأ ويسمى اسمها، وينصب الخبر ويسمى خبرها. ومن شروطها حذف اسمها، وأن يكون اسمها وخبرها من أسماء الزمان. [٢]

- **لا (النافية) LAA**

حرف معنى مبني على السكون لا محل له من الإعراب، ويشمل هذا الحرف: حرف النفي المشبه بليس (لا شجاعٌ مخذولا)، حرف النفي للجنس (لا رجل في الدار)، حرف النهي (لا تفعل ذلك)، حرف جواب (أمسافر أنت...لا) ، حرف النفي (لا يوجد عمل هنا). [٣]

- **غير عاملة NoSyntaEffect**

حرف معنى لا عمل له، أي لا يحدث تغييرًا شكليًّا في غيره سواء أكان اسما أو فعلا، ومن هذه الحروف غير العاملة: قد، كلا، سوف، كلا، بلى، نعم، إذا، إنْ، ما، أمّا، ربما،

---

(١) المصدر السابق، ص٨.

(٢) المصدر السابق، ص٢٧٤.

(٣) المصدر السابق، ص٢٧٢.

لولا، آنذاك، لو، إي، هلا. [1]

- **ناصب ParticleNAASSIB**

هي الأداة التي تنصب الفعل المضارع، مثل: أن، لن، إذن، كي أو لكي، حتى، أن المضمرة.

- **مصدرية MASSDARIYYA**

يقصد بالمصدرية إمكان تأويل حرف "ما" المصدري مع ما بعده بمصدر، ويطّرِد وقوعها في الكلمات الآتية: بينما، حينما، بعدما، طالما، مثلما،...

٣-١-٣-١ خصائص الأداة الصرفية

٣-١-٣-١-١ علامات اللواصق:

- **لا سابق NullPrefix**

هي علامة تستخدم للدلالة على عدم وجود سابق، يسبق الأداة.

- **لا لاحق NullSuffix**

هي علامة تستخدم للدلالة على عدم وجود لاحق، يلحق الأداة.

ومن أمثلة عنونة القسم الكلامي "الأداة" في المدونة اللغوية:

الكلمة	العنونة باللغة العربية	العنونة باللغة الإنجليزية
إلى	لا سابق+حرف جر+لا لاحق	NullPrefix+Prepos+NullSuffix
يا	لا سابق+حرف نداء+لا لاحق	NullPrefix+Interj+NullSuffix
لها	لا سابق+جار ومجرور+لا لاحق	NullPrefix+PrepPronComp+NullSuffix
إن	لا سابق+حرف ناسخ+لا لاحق	NullPrefix+ParticleNAASIKH+NullSuffix
أو	لا سابق+حرف عطف+لا لاحق	NullPrefix+Conj+NullSuffix
لقد	لام توكيد+ غير عاملة+لا لاحق	Confirm+NoSyntaEffect+NullSuffix

(١) المصدر السابق، ص٢٧٤.

الكلمة	العنونة باللغة العربية	العنونة باللغة الإنجليزية
هل	لا سابق+أداة استفهام+لا لاحق	NullPrefix+InterrogArticle+NullSuffix
لم	لا سابق+جازمة+لا لاحق	NullPrefix+JAAZIMA+NullSuffix_CON
عندما	لا سابق+ شرطية جازمة+ظرف+لا لاحق	NullPrefix+CondJAAZIMA+DZARF+NullSuffix_CON
لو	لا سابق+ شرطية غير جازمة+ظرف+لا لاحق	NullPrefix+CondNotJAAZIMA+NoSyntaEffect+NullSuffix_CON
إلا	لا سابق+استثناء+لا لاحق	NullPrefix+Except+NullSuffix
لا	لا سابق+لا+لا لاحق	NullPrefix+LAA+NullSuffix
أما	لا سابق+غير عاملة+لا لاحق	NullPrefix+NoSyntaEffect+NullSuffix
كي	لا سابق+ناصب+لا لاحق	NullPrefix+ParticleNAASSIB+NullSuffix
بينما	لا سابق+ظرف+مصدرية+لا لاحق	NullPrefix+DZARF+MASSDARIYYA+NullSuffix

الجدول:٢٥ عنونة القسم الكلامي " الأداة" في المدونة اللغوية عينة الدراسة

٣-١-٤ الظرف

هو كلمة تدل على الزمان أو المكان الذي يقع فيه الفعل، ويتضمن معنى "في" باطراد. وهناك نوعان من الظروف، ظرفٌ غير متصرف لا يستعمل إلا ظرفًا، وظرفٌ متصرف يستعمل ظرفًا وغير ظرف. وما يعنينا هنا الظرف الخالص فحسب.[١]

- **الظرف DZARF**

تعبّر هذه العلامة عن الظرف غير المتصرف الذي لا يتسعمل إلا ظرفًا مثل: حيث، إذ، قط، لدن، ثم، أمام، عند، دون، مع، لدى.

(١) المصدر السابق، ص١٩٢.

٣-١-٤-١ خصائص الظرف الصرفية

٣-١-٤-١-١علامات اللواصق:

- **لاسابق NullPrefix**

هي علامة تستخدم للدلالة على عدم وجود سابق يسبق الظرف.

- **لا لاحق NullSuffix**

هي علامة تستخدم للدلالة على عدم وجود لاحق يلحق الظرف.

ومن أمثلة عنونة القسم الكلامي " الظرف" في المدونة اللغوية:

الكلمة	العنونة باللغة العربية	العنونة باللغة الإنجليزية
إذ	لا سابق+ظرف+لا لاحق	NullPrefix+DZARF+NullSuffix
هنا	لا سابق+ظرف+لا لاحق	NullPrefix+DZARF+NullSuffix
تحت	لا سابق+ظرف+مضاف+لا لاحق	NullPrefix+DZARF+Adjunct+NullSuffix
عند	لا سابق+ظرف+مضاف+لا لاحق	NullPrefix+DZARF+Adjunct+NullSuffix

الجدول:٢٦ عنونة القسم الكلامي " الظرف" في المدونة اللغوية عينة الدراسة

٣-١-٥ كلمة غير عربية Translit

هي كلمة معربة تعريبًا صوتيًّا، وتشيع في العربية المعاصرة لا سيما في لغة الصحافة. مثال ذلك تكنوقراطي، سايكس-بيكو، واشنطن، اليونيفل.

ومن أمثلة عنونة القسم الكلامي " كلمة غير عربية" في المدونة اللغوية:

الكلمة	العنونة باللغة العربية	العنونة باللغة الإنجليزية
أثينا	كلمة غير عربية	Translit
اسلام_اونلاين	كلمة غير عربية	Translit
الأسكواش	كلمة غير عربية	Translit
الريموت	كلمة غير عربية	Translit
السينما	كلمة غير عربية	Translit
الكلور	كلمة غير عربية	Translit
الكمبيوتر	كلمة غير عربية	Translit
النيوترون	كلمة غير عربية	Translit
أنفلونزا	كلمة غير عربية	Translit
جاكلين	كلمة غير عربية	Translit
جورج	كلمة غير عربية	Translit
ديزني	كلمة غير عربية	Translit
سانت_أندروز	كلمة غير عربية	Translit

الجدول:٢٧ عنونة القسم الكلامي ” كلمة غير عربية“ في المدونة اللغوية عينة الدراسة

٣-١-٦ الترقيم Punc

هو مجموعة من العلامات الرمزية التي تستخدم لتنظيم النص الكتابي، بحيث يسهل تقسيمه وقراءته. وهي الفاصلة (،)، الفاصلة المنقوطة (؛)،النقطتان الرأسيتان (:) علامة الاستفهام(؟)، علامة التعجب (!)، الشرطتان (- -)، القوسان ( )، علامة الحذف (...)، علامة التنصيص (“”)، علامة الخط المائل (/).

ومن أمثلة عنونة القسم الكلامي ” الترقيم“ في المدونة اللغوية:

الكلمة	العنونة باللغة العربية	العنونة باللغة الإنجليزية
،	علامة ترقيم	PUNC
؛	علامة ترقيم	PUNC
.	علامة ترقيم	PUNC
:	علامة ترقيم	PUNC
"	علامة ترقيم	PUNC
؟	علامة ترقيم	PUNC
!	علامة ترقيم	PUNC
/	علامة ترقيم	PUNC
-	علامة ترقيم	PUNC
(	علامة ترقيم	PUNC
)	علامة ترقيم	PUNC
...	علامة ترقيم	PUNC
'	علامة ترقيم	PUNC

الجدول:٢٨ عنونة القسم الكلامي " الترقيم" في المدونة اللغوية عينة الدراسة

•••

الفصل الثالث

العنونة بالعلاقات التركيبية

مقدمة:

تأتي العنونة بالعلاقات التركيبية في قمة الهرم التدريجي للبناء، فبها يكتمل البناء الشجري النحوي الذي تمخض عن تكامل المراحل الثلاث التنظيمية (المدونة اللغوية - العنونة بالأقسام الكلامية - العنونة بالعلاقات التركيبية). وقد تناولت في هذا الفصل -بداية- تعريف العلاقات التركيبية، ودورها في تحديد أبنية الجمل، ثم شرعت في بيان مفهوم العنونة بالعلاقات التركيبية، وأهميتها على الصعيدين اللغوي والحاسوبي، ثم عرضت الطرق الإجرائية التي تنطلق منها العنونة التركيبية (التمثيل النحوي - المحتوى النحوي - النظرية النحوية)، وأخيرًا يعرض الفصل نماذج من عنونة المدونة اللغوية - عينة الدراسة - بالعنونة بالعلاقات التركيبية على المستوى المكوناتي والوظيفي. وقد جاء هذا الفصل في ثلاثة مباحث على النحو التالي:

١. **المبحث الأول: العلاقات التركيبية.**
- ما أهمية التحليل بالعلاقات التركيبية؟

٢. **المبحث الثاني: العنونة بالعلاقات التركيبية.**
- لماذا العنونة بالعلاقات التركيبية؟
- التمثيل النحوي.

- المحتوى النحوي.
- النظرية النحوية.

٣. المبحث الثالث: عنونة المدونة اللغوية - عينة الدراسة - بالعلاقات التركيبية.

- فئة العناصر الأساسية للتحليل التركيبي.
- فئة السمات النحوية الوظيفية.

المبحث الأول

العلاقات التركيبية

١. العلاقات التركيبية

يهدف علم التركيب Syntax إلى إيضاح القوانين والعلاقات التي تحكم أبنية الجمل، لمحاولة فهم النظام الذي يجري عليه الائتلاف بين مكونات الجمل، إذ التعرف على العلاقات التركيبية المتمثلة في الربط والارتباط بين أجزاء الجملة، يؤدي إلى إبراز المعنى الدلالي المستهدف من ذلك النظم.

ويرتبط مفهوم التركيب في الجملة وعناصرها والعلاقة الناشئة بين وحداتها بالدراسات اللغوية الحديثة بيد أنه متأصل في تراثنا النحوي، حيث توصل عبد القاهر الجرجاني في كتابه دلائل الإعجاز إلى نظرية تركيبية محكمة ألا وهي نظرية التعليق، تنطلق في دراسة الجملة من المعنى وصولًا إلى المبني؛ أي وفق ما تسير عليه عملية الاتصال اللغوي. ويرسي عبد القاهر دعائم هذه النظرية بمبدئه الذي حدد أبعاد الجملة من الناحية الدلالية، إذ يقول: «إذا قلت: ضرب زيدٌ عمرًا يوم الجمعة ضربًا شديدًا تأديبًا له. فإنك تحصل مجموع هذه الكلم كلها على مفهوم، هو معنى واحد لا عدة معان، كما يتوهمه الناس»[١]. وهذا يرشدنا إلى أن الجملة ذات معنى دلالي واحد، ينشأ

(١) الجرجاني (عبد القاهر): دلائل الإعجاز، تحقيق أبي فهر محمود محمد شاكر، مكتبة الخانجي - القاهرة، ط١، ١٩٩١م، ص٤١٣.

من التفاعل بين المعاني الجزئية المتمثلة في مكونات الجملة أي أن التركيب هو الوسيلة المباشرة لنشوء هذا المعنى الواحد.

ومع ذلك لم نحاول استغلال هذه النظرية في درسنا النحوي ولا توظيفها في موضع ملائم لها،[١] هكذا تعودنا - للأسف - على أن ننتظر أن تأتي لنا الحلول جاهزة من قبل العلماء الغربيين في ظل توطد العلاقة بين اللغة والعلوم الأخرى التي تهدف إلى فهم النظام اللغوي.

اتجه علماء اللغة الغربيون في الأربعينات من القرن الماضي إلى الوصف والتنظير للجملة بشكل عام في إطار العديد من المنطلقات الفكرية والفلسفية؛ ليخرجوا بمسلمة مفادها أن محور تحليل الجملة هو العلاقات التركيبية.[٢]

ثم أثمر التفاعل بين المستويات اللغوية المختلفة عدة نظريات لغوية تنطلق جميعها من العلاقات التركيبية باعتبارها قالبًا تنتظم المعاني بداخلها في إطار مبادئ النظرية، على سبيل المثال نظرية دي سوسير التي اتخذت العلاقات وسيلة لها في دراسة التفاعل بين الكلمات وتأثير ذلك في تكوين الجمل. وكذلك نظرية بلومفيلد التوزيعية التي اعتمدت على تجزئة الجملة إلى مكوناتها المباشرة ثم إلى مكوناتها النهائية بهدف توزيع الوحدات اللغوية، مع أنها تنتحي بالمعنى جانبا... ولعل ذلك يدفعنا إلى التساؤل هل كشف بلومفيلد عن آلية محكمة في تقسيم المكونات المباشرة للجملة في ظل استبعاد المعنى من التحليل؟

ثم اتسعت دائرة التفاعل بين اللغة والعلوم الأخرى لاحتواء إشكالية اللغة المتشعبة، الأمر الذي أدى إلى انطلاق العديد من النظريات النحوية العامة؛ لتفسير الظواهر التركيبية في إطار رياضي ومنطقي.

وقد تمركز جل هذه النظريات في اتجاهين أساسيين لتحليل الظاهرة النحوية في إطار

(١) يقف الدكتور مصطفى حميدة موقف المتعجب من هذا الجمود الفكري وعدم التجديد في إطار هذه النظرية التي تقف اليوم شامخة أمام العديد من النظريات اللغوية الحديثة، إذ يقول: «ولعل أهم ما أسداه عبد القاهر للعربية أنه وضع الدرس النحوي في طريقه الصحيح، وحدد له المنهج الذي يشمله من جميع جوانبه، فكان من المأمول أن يبدأ الدرس النحوي مرحلة جديدة من التطور في ضوء تلك النظرية المتكاملة، لكن الأمر الغريب أن من جاؤوا بعد عبد القاهر لم يفهموا مرساه من كتابه دلائل الإعجاز، وكان للصناعة النحوية سلطان عظيم في نفوسهم، وكان منهج الإعراب قد رسخ في العقول رسوخا، ولم يقع في أخلادهم أن دلائل الإعجاز كتاب في النحو، وأن النظرية التي عرضها هي أساس الدرس النحوي وذروة فلسفته ومنهجه القويم، فحين وجدوه يتحدث عن المعاني وترتيبها في النفس وملاءمة معنى اللفظة لمعنى التي تليها واتحاد أجزاء الكلام ووضع الجملة في النفس وضعا واحدا ظنوه يتحدث في علم جديد لايمت بصلة إلى علم النحو الذي ألفوه، فأطلقوا على ذلك العلم علم المعاني وجعلوه مبحثا في علم البلاغة.» ينظر:
حميدة، (مصطفى): الربط والارتباط في تركيب الجملة العربية، مكتبة لبنان ناشرون، ط١، ١٩٩٧، ص ١٣٠.

(٢) عبادة (محمد إبراهيم): الجملة العربية، مكوناتها - أنواعها - تحليلها، مكتبة الآداب-القاهرة، ط ٢٠٠٧، ص٦.

العلاقات التركيبية، وهما الاتجاه التوليدي الذي يصوغ تلك العلاقات في مكونات تركيبية، والاتجاه التحليلي الذي يصوغ تلك العلاقات في اعتمادات مفرداتية.

١-١ ما أهمية التحليل بالعلاقات التركيبية؟

التساؤل عن أهمية التحليل بالعلاقات التركيبية دائم، واختلاف مشاغل القائمين عليه حادث حسب اختلاف المادة النحوية المدروسة[١]، وقد لخص ابن خلدون في مقدمته تلك الأهمية، فيقول: «به تتبين أصول المقاصد بالدلالة فيعرف الفاعل من المفعول والمبتدأ من الخبر، ولولاه لجهل أصل الإفادة.» [٢] كما أن له أهمية في تفسير دلالة النص، ومعرفة الخصائص الأسلوبية للنص النثري والشعري، وتحليل الخطاب؛ ومن ثم استنباط القوانين التي تحكمه.

٢. العنونة بالعلاقات التركيبية

هي عملية تعيين العلاقات النحوية التي تنشأ بين أجزاء الجملة؛ نتيجة حدوث علاقات نحوية سياقية بين المعاني، وذلك في إطار نموذج نحوي مصطبغ بشكل رياضي (بنية العبارة أو بنية التبعية)، تمهيدًا للمعالجة الآلية للمنظومة النحوية.

ويتم تمثيل تلك العلاقات لجمل المدونة اللغوية باستخدام مجموعة من الرموز Tag set (فاعل، مفعول، إضافة، وصف، ...) للتعبير عن المحتوى النحوي المراد إضافته للمدونة اللغوية.

٢-١ لماذا العنونة بالعلاقات التركيبية؟

تمثل المدونات المعنونة بالعلاقات التركيبية موردًا هامًّا لمعالجة النحو آليًّا، «تشهد ساحتها أي المعالجة أقصى درجات الامتزاج بين اللسانيات والحاسوبيات»[٣]، كما تعدُّ منطلقا للتحقُّق من فرضيات النَّظريَّات اللُّغوية[٤]، فضلًا عن رسم معالم واضحةٍ لنظام

---

(١) عاشور (المنصف): بنية الجملة العربية بين التحليل والنظرية، ، ص ٧.

(٢) ابن خلدون (ولي الدين عبد الرحمن بن محمد): مقدمة ابن خلدون، تحقيق عبد الله محمد الدرويش، دار البلخي - دمشق، ط١، الجزء الثاني، ٢٠٠٤، ص ٣٦٧.

(٣) علي (نبيل): اللغة العربية والحاسوب، ص ٣٨٨.

(4) Abeillé, A. (2003). Building and Using Parsed Corpora, Springer. P.XV.

الجملة في اللُّغة المَدروسة وتحديد خصائص علاقاتها التركيبية، مبيِّنة وجوه الائتلاف والاختلاف، والتَّغَيُّر أو عدم التَّغَيُّر في بنية الجملة.[١] وتتلخص أهداف المدونات الموصَّفة بالعلاقات التركيبية في هدفين أساسيين:

• **بناء المحللات النحوية**

يعدّ السبب الرئيسي لعنونة المدونات بالعلاقات التركيبية هو استخدامها في تدريب خوارزميات التعلم الآلي لبناء المحللات النحوية[٢]؛ وذلك للاستفادة منها في تطبيقات معالجة اللغة الطبيعية المختلفة مثل: الترجمة الآلية، إجابة الأسئلة، تعلّم النحو، التدقيق النحوي، تحديد أساليب الكتَّاب ومعرفة هويتهم، التخاطب مع قواعد البيانات باللغة الطبيعية، التخاطب مع الإنسان الآلي Robots الذي يتم عن طريق التعرف على الكلام ثم تحليله نحويًّا؛ حتى يتحصَّل على المقاصد الدلالية.

وما زالت المحللات النحوية إلى الآن – مذ ما يربو على عشرين عامًا في البحث والتطوير - تفتقر إلى الصلابة، واتساع التغطية التي تحتاجها في تحليل النصوص المتنوعة أو الخطاب.[٣] وربما يعود ذلك لتنوع المستوى اللغوي - من حيث الفصاحة والعامية - في ضوء بزوغ مجتمع المعرفة.

• **الدراسات اللغوية**

يمكن أن تكون المدوَّنة الموصَّفة نحويًّا مصدرًا ثريًّا للمعلومات التي تستخدم في بناء العديد من التطبيقات، مثل استخلاص الإطار النحوي للأفعال Subcategorization Frames of Verbs التي تحدد عن طريقة بنية العوامل Arguments Structure،[٤] واستخلاص الأنماط التركيبية في المدونة اللغوية، واستخلاص متوسط أطوال الجمل، واستخلاص الجمل التكميلية Relative clauses ومتوسط أطوالها، واستخلاص التراكيب النحوية بأنواعها المختلفة، واستخلاص الأفعال المتعدية إلى مفعولاتها المباشرة أو غير المباشرة، واستخلاص الخصائص التركيبية، ومن ثم وصف نظام اللغة آنيًّا وزمانيًّا وسياقيًّا،... وغيرها.

(١) عاشور (المنصف): بنية الجملة العربية بين التحليل والنظرية، ص٨.

(2) Habash, N. (2010). Ibid. P.104.

(3) Garside, R. & Leech, G. & McEnery, T. (1997). Ibid.36.

(4) Garside, R. & Leech, G. & McEnery, T. (1997). Ibid.34-35.

٣. الطرق الإجرائية للعنونة بالعلاقات التركيبية

وتتوخى العنونة بالعلاقات التركيبية عدة طرق إجرائية، لا تنفك إحداهن عن الأخرى، فهي بمثابة أجزاء اللوحة التشكيلية التي لا يكتمل معناها إلا إذا اتحدت وتكاملت مع بعضها البعض[١]. وهذه الطرق المتوخاة متأصلة في التحليل النحوي الحديث إلا أن الخلط بينها دائم ربما لإغفال الجانب التطبيقي الملموس في الدرس النحوي. وتتحقق العنونة بتوافر تلك الطرق المتلازمة الآتية:

٣-١ التمثيل النحوي Syntactic Representation

المتتبع لتراكيب الجملة في أي لغةٍ سيلاحظ أن اللغة تخضع لمنظومة قاعدية في الذهن البشري، ربما تكون مُنشأة تمامًا وواحدة البنية في كل الأدمغة، بحيث إن إنتاج الجملة في أي لغة يخضع لنمط معين وترتيب خاص يفهمه المنتمون للمجموعة اللسانية نفسِها، وأي خلل في المواضع يؤدي فورا إلى فقد التواصل أو اضطرابه، وهو ما دفع تشومسكي إلى القول بأن النحو الكلي Universal Grammar (UG) [٢] موجود في أذهان الأطفال منذ الولادة، وهو أساس كل اللغات التي يتعلمها البشر، ويتمثل بكونه مجموعةً من المبادئ والأحياز (المحددات) Principles and Parameters، واستجابةً لما يوجد في المحيط يبتكر الطفل نحوا مشتركا يَسِمُ كل هذه الأحياز بقيم Values مُنتِجًا إحدى لغات البشر الممكنة التي يبدأ بها، بحيث إن ذهن الطفل ينفتح أول مرة إلى أي لغة بشرية، منتهيا إلى اكتساب لغة معينة.[٣]

وأخيرًا أجمع باحثو اللسانيات العصبية NeuroLinguistics – من خلال التجارب – أن المنطقة اليسرى من الفص الصدغي الأمامي Left anterior Temporal Lobe تطبق نوعًا

(1) see: Rambow, O. (2010). The Simple Truth about Dependency and Phrase Structure Representations.Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the ACL, Los Angeles, California.

(٢) إذا نظرنا إلى اللغات الاصطناعية (البرمجية) عند تعاملها مع الآلة، سنجد أنها لا تتعامل مباشرة مع الآلة؛ بل هناك وسيط compiler يترجمها – أي اللغات البرمجية – إلى لغة Assembly التي تفهمها الآلة؛ لأنها – ببساطة – تقوم على رقمي الصفر والواحد. وهذا يعني أن هناك بنية لغوية واحدة تشترك فيها جميع اللغات البرمجية باختلاف تراكيبها، وهذا ما يتوافق نسبيًّا مع عمل اللغة بشكل عام، ويدعم فكرة البنية اللغوية المشتركة.

(٣) مزوني (دومينيك): قيود النحو الكلي: كيف يكتسب البشر القدرة على الكلام، ترجمة فرحات المليح. مجلة الحياة الثقافية. السنة ٢٧. العدد ١٣٩، ٢٠٠٢، ص ٣٦.

ما من المعالجة النحوية الأساسية[1]، وهذا ما يدعم مسألة وجود نمط معين مُنشأ بالدماغ البشري، إلا أنهم اختلفوا في تفسير نوع هذه المعالجة، حيث ما زال يكتنفها الغموض.

ومع ذلك يجتهد علماء اللغة محاولين تمثيل تلك المعالجة النحوية الموجودة في الذهن البشري من خلال البناء الهندسي للفضاء الذهني وتصور المعنى في الدماغ الذي ينطلق من مسلمة ذهنية مفادها ”أن المعنى في اللغة الطبيعية بنية معلومات مرمزة في الذهن البشري أو هو تمثيل ذهني، ومن ثمة فإن المعلومات التي تحملها اللغة مصوغة بالطريقة التي ينظم بها الذهن التجربة، ولا يمكن لهذه المعلومات المتجلية في تعابير البنية التصورية أن تحيل إلى العالم الواقعي، كما في نظريات أخرى، وإنما إلى عالم مُسقط ناتج من هذه البنية، ووليد التنظيم الذهني المذكور“.[2]

وانطلاقا من هذا الإطار التصويري لبنية المعنى داخل الذهن، افترض التوليديون والتحليليون وغيرهم من علماء اللغة أصحاب النظريات النحوية الحديثة نموذجًا افتراضيًّا لشكل المعرفة النحوية في ذلك الفص الصدغي، يتمثل في أن المعرفة النحوية عبارة عن وحدات مترابطة أو متداخلة تتفرّع عن بعضها البعض.[3] وتأتّى ذلك لهم مقاربةً من شكل المشتبكات العصبية synopses في الخلايا العصبية المسؤولة عن نقل الإشارات الكهربائية -التي تحمل المعلومات - بين تلك الخلايا.[4]

ثم حاول هؤلاء العلماء تطبيق القوانين الرياضية والمنطقية لنقل أنساق هذا النموذج المعرفي الذهني إلى أنساق صورية؛ لمحاولة تلمس هذا العالم الخفي وإدراك البنية العقلية اللغوية، فانطلقوا جميعا من مسلمة مفادها أن النموذج النحوي الافتراضي يتمثل في عدد محدود من العلاقات والقيود، تتدرَّج فيما بينها في هيكل تنظيمي أي شجري.[5]

وقد أثمر المنطق الرياضي في صوغه للنموذج النحوي تمثيلين نحويين، تعددت في إطارهما العديد من النظريات النحوية الحديثة، وهما:

---

(1) Hale, J. & Callaway, F. & Feldman, E. and Others. (2014). Modeling Neural Correlates of Syntactic Structure Building. In AMLaP 2014-Poster, Scotland.

(٢) غاليم (محمد): هندسة التوازي النحوي وبنية الذهن المعرفية. كتاب آفاق اللسانيات (تكريما للأستاذ الدكتور نهاد الموسى)، مركز دراسات الوحدة العربية، بيروت، ط ١، ٢٠١١، ص ٥٢.

(3) Dirven, R. & Langacker, R. (1992). Grammar in Mind and Brain. Mouton de Gruyter, Berlin. New York. P.251 and 277.

(٤) شريف (عمرو): ثم صار المخ عقلا، طبعة مكتبة الشروق الدولية، ط ٢، ٢٠١٣، ص ٥٥-٥٦.

(٥) لعل هذا الطرح يهدينا إلى إجابة مقنعة عن السؤال المحيِّر الذي يطرحه العديد من الباحثين، وهو ما فائدة التشجير في التحليل النحوي؟ بالإضافة إلى أنه الأقدر على تمثيل تلك المعرفة النحوية من الناحية الرياضية المنطقية.

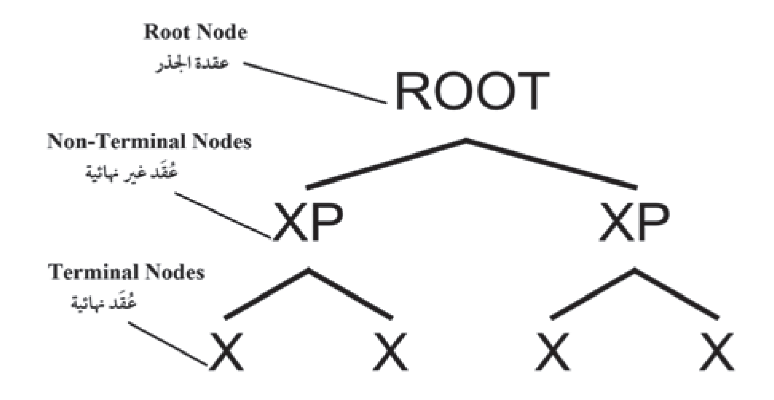
- **التمثيل المكوني Constituency Representation**
- **التمثيل الاعتمادي Dependency Representation**

وبناءً على ذلك يمكن تعريف **التمثيل النحوي** بأنه النموذج الرياضي الذي يعرض بنية الجملة بشكل تصويري في إطار النظرية النحوية والمحتوى النحوي. وقد أسهم هذا التمثيل النحوي في توضيح طبيعة المعرفة وأنساقها والفهم والتأويل، وفي التقدّم التقني للحوسبة computation.[1]

٣-١-١ التمثيل المكوني Constituency Representation

هو بنية شجرية منظمة، تنتظم فيها كلمات الجملة في شكل مركبات أو مكونات نحوية تبعًا لنظامٍ نحويٍّ، بحيث تظهر فيها الكلمات كعُقَد نهائية Terminal والمركبات كعُقَد غير نهائية Non-Terminal. وهو ما يعرف بـ أشجار بنية العبارةPhrase-Structure Trees .

ويمكن توضيح العقد Nodes في بنيتها الشجرية من خلال الشكل رقم ٢٧:



الشكل:٢٧ مخطط هيكلي لبنية الشجرة النحوية

حيث X تعني الكلمات أو الوحدات، بينما XP (X Phrase) تعني المركبات أو المكونات النحوية (مركب اسمي، مركب فعلي، مركب حرفي،...).

ويتم تنظيم هذه البنية الشجرية أو كتابتها عن طريق:

(١) الفهري (عبد القادر الفاسي). البناء الموازي، نظرية في بناء الكلمة وبناء الجملة، دار توبقال للناشر، المغرب، ط ١، ١٩٩٠، ص ١٧.

٣-١-١-١ التقويس Bracketing

هي عملية رياضية تنظم كلمات الجملة ومركباتها بطريقة متماسكة، بحيث تظهر العلاقات بينهما في صورة اعتمادية [١] وقد اعتمدها عالم اللغويات الحديثة نعوم تشومسكي في تمثيله لقواعد النحو المتحرر من السياق.[٢]

إن إطلاع تشومسكي على العديد من العلوم الصورية Formal Science لاسيما الرياضيات والمنطق وعلوم الحاسوب، واشتغاله بمشروع الترجمة الآلية بمعهد ماساشوستس للتقنية Massachusetts Institute of Technology (MIT) في الخمسينات من القرن المنصرم [٣]، جعله ينظر للغة على أنها منظومة متسقة البناء، شديدة الإحكام، يمكن أن تمثل في إطار رياضيّ حاسوبيّ؛ لذا نجد أثر هذا التفاعل في كتابه التراكيب النحوية عام ١٩٥٧م، حيث اقترح ثلاث طرق للتحليل النحوي، تنطلق جميعها من المنطق الرياضي الحاسوبي.

ففي الطريقة الأولى استخدم تشومسكي آلية الحالة المحدودة Finite-State Machine (FSM) – التي تستخدم في تصميم التطبيقات الهندسية والبرمجيات الحاسوبية[٤] – لتوليد الجمل بشكل تتابعي اعتمادًا على الحالة التي تحدد العناصر التالية لها. ومع ذلك "قد اختبر تشومسكي هذه الطريقة ووجدها غير صالحة للتحليل اللغوي؛ لأنها تؤدي إلى تقديم جمل محدودة، بينما تقدم اللغة جملا لا نهاية لها".[٥]

أما الطريقة الثانية والثالثة فاستخدم تشومسكي النظام الرياضي والمنطق الرمزي؛ لوصف المكونات المباشرة حسب عناصرها في إطار القوانين والأحكام، وذلك عن طريق أقواس التحليل Analysis Brackets التي تتشابه مع أقواس الرياضيات وأساسيات العديد من اللغات البرمجية مثل لغة ليسب lisp التي تنظم بياناتها وقوائمها بالتقويس Bracketing التحليلي. ويمكننا أن نبيّن مدى الإحكام في طريقة التقويس التحليلي بالتمثيل الرياضي:

(1) Pustejovsky, J & Stubbs, A. (2013). Ibid. 18.

(2) Jurafsky, D. & Martin, J. (2007). Ibid. P.433.

(3) Otero, C. (1994). Noam Chomsky: Critical Assessments.TJ Press (Padstow) Ltd, Padstow, Cornwall, London, P.433.

(4) Wright, David R. (2005). "Finite State Machines". CSC215 Class Notes. Prof. David R. Wright website, N. Carolina State Univ. Retrieved July 14, 2012.

(٥) الراجحي (عبده). النحو العربي والدرس الحديث، بحث في المنهج، دار النهضة العربية – بيروت، ط١، ١٩٧٩م، ص ١٣٢.

فمثلا إذا كان عندنا هذه المعادلة:

س(ص+ع)

وفقا لأسبقية ترتيب العمليات الحسابية في الرياضيات أو البرمجيات الحاسوبية، نجد أن مُعامل العمليات المدمجة داخل الأقواس يتصدر تلك العمليات الحسابية ثم يليها الضرب المتكرر، ثم الضرب والقسمة ثم الجمع والطرح... وبناءً على ذلك فإن أي تغيير في العمليات الحسابية في هذه المعادلة سيؤدي إلى اختلاف في النتائج.

لنفترض أن:

س = ٢   ص= ٤   ع= ٦

وبالتعويض في المعادلة: س(ص+ع)

= ٢ (٤+٦)

= ٢ × ١٠

إذن مجموع المعادلة = ٢٠

في حين انتزاع الأقواس ستصبح المعادلة هكذا: س× ص+ع

وبالتعويض في المعادلة: س× ص+ع

=٢ ×٤+٦

إذن مجموع المعادلة = ١٤

وكذلك يتم الإحكام في التمثيل النحوي للجمل عن طريق التقويس في إطار المعلومات النحوية المقدمة.

ويوضح المثال التالي مدى إحكام الأقواس الرياضية في تمثيل المعاني النحوية:

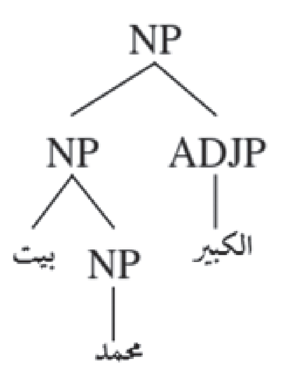
مثلا جملة: بيت محمد الكبير ...فإن كلمة " الكبير" تحتمل أن تكون صفة لـ محمد أو صفة لـ البيت. ومن خلال الأقواس يتبيَّن المعنيان:

١- المعنى الأول، صفة "الكبير" عائدة على البيت.

```
(NP
    (NP ( بيت) (NP ( محمد)))
    (ADJP ( الكبير)))
```

وبتحويلها للشكل الشجري[1] :

(١) يتم تحويلها آليًا - باستخدام العديد من الأدوات والحزم البرجية مثل stanford-tregex Const Tree & Viewer – من خلال تنفيذ عمليات ترتيب الأقواس. وقد اعتمد تشومسكي على هذا النظام أي التقويس، ونظرًا لصعوبته في التحليل، فقد اعتمد أتباعه على النظام المحول له أي الشجري. انظر: اللغة والمسؤولية د.حسام البهنساوي ص ٦٧.

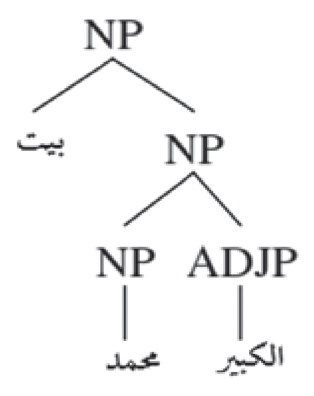


فإذا نظرنا إلى مركب الصفة ADJP "الكبير" سنجده موازيًا لرأس المركب الاسمي NP "البيت" من ناحية السيطرة المباشرة Immediately dominates من المركب الاسمي NP الأكبر. وهو ما يعبر عن التوازي النحوي بين المركبين.

المعنى الثاني، صفة "الكبير" عائدة على محمد.

```
(NP
 بيت
  (NP (NP محمد) (ADJP الكبير)))
```

وبتحويلها للشكل الشجري :



فإذا نظرنا إلى مركب الصفة ADJP "الكبير" سنجده موازيا للمركب الاسمي NP "محمد" من ناحية السيطرة المباشرة Immediately dominates من المركب الاسمي NP المكمل للمركب الاسمي الأكبر. وهو ما يعبر عن التوازي النحوي بين المركبين.

٣-١-١-١-١ خصائص التقويس في التمثيل النحوي:

- يعد التقويس هو الطريقة الأساسية لتمثيل أشجار بنية العبارة حاسوبيًّا.
- يصعب قراءته على غير المتمرن عليه، لكن يمكن تحويله لصورة – مرئية – شجرية من خلال بعض الأدوات الحاسوبية.
- إمكانية ضبط التحليل الشجري في شكل متدرّج، بحيث يكشف عن العلاقات بين المكونات في صورة هرمية كما في الأمثلة السابقة.
- إمكانية التعديل والحذف والاستبدال في البنية الشجرية عن طريق استخدام التعبيرات النمطية Regular Expressions .
- إمكانية استخلاص المعلومات النحوية أو الأشجار الفرعية Sub-trees من البنية الشجرية.

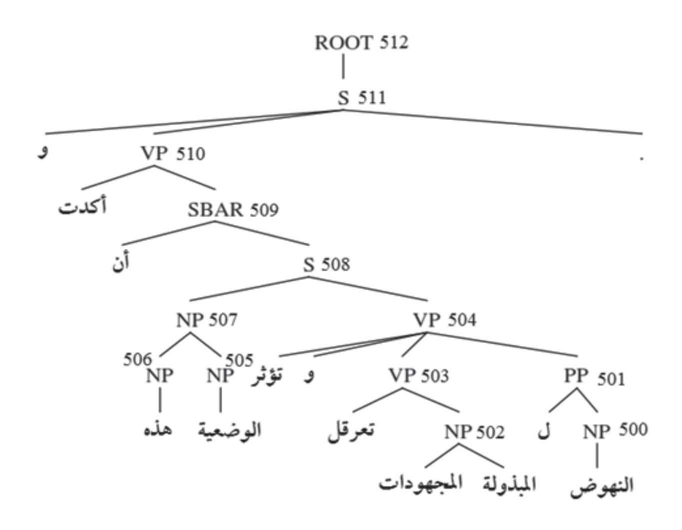
٣-١-١-٢ هيكلة نيجرا NEGRA Format

هي طريقة وصفية رياضية، تنظم البنية الشجرية للجملة في صفوف وأعمدة ذي علاقات ترابطية، حيث تخزن الكلمات وعناوين المركبات Tags of Phrases في تلك الأعمدة بطريقة رقمية ترتيبية، يمكن قراءتها بسهولة. وقد اعتمد أستاذ اللغويات الحاسوبية ثورستن برانتس Thorsten Brants بجامعة سارلاند الألمانية Saarland University على هذه الطريقة في وصف المدونات المعنونة Annotated Corpora بالعلاقات التركيبية لا سيما مدونة نيجرا النحوية Syntactic NEGRA.[(١)]

وكان الهدف من هذه العملية وصف المدونات المعنونة بالأقواس النحوية لا سيما مدونات بنسلفانيا الموصَّفة؛ حتي يتسنى قراءتها، ومن ثم إمكانية تحريرها بسهولة.

ويوضح المثال التالي كيفية قراءة مكونات البنية الشجرية المكونية؛ لتحويلها إلى معلوماتٍ لغوية في صفوف وأعمدة.

(1) Brants, T. (1997). The NeGra Export For Annotated Corpora. Projekt C3 NebenLaufige Grammatische Verabeitung.



الشكل:٢٨ البنية الشجرية المكونية

ومن خلال قراءة المركبات من الأسفل إلى الأعلى في الشجرة السابقة، يمكننا رصدها في صورة جدول:

#BOT

Word ID	Word	Phrases	Node ID
S1_1	و	_ _	511
S1_2	أكدت	_ _	510
S1_3	أن	_ _	509
S1_4	هذه	_ _	506
S1_5	الوضعية	_ _	505
S1_6	تؤثر	_ _	504
S1_7	و	_ _	504
S1_8	تعرقل	_ _	503
S1_9	المجهودات	_ _	502
S1_10	المبذولة	_ _	502
S1_11	ل	_ _	501
S1_12	النهوض	_ _	500
S1_13	.	_ _	511
	#500	NP	501
	#501	PP	504
	#502	NP	503

#503	VP	504
#504	VP	508
#505	NP	507
#506	NP	507
#507	NP	508
#508	S	509
#509	SBAR	510
#510	VP	511
#511	S	500
#512	ROOT	0

#EOT

الجدول:٢٩ تمثيل هيكلة نيجرا

تبدأ العنونة بالكلمة المفتاحية "بداية الجدول" BOT# (Beginning Of Table)، وتنتهي بالكلمة المفتاحية "نهاية الجدول" EOT# (End Of Table)، ويتوسطهما المعلومات التحليلية للجملة في صفوف وأعمدة، فالعمود الأول يعبر عن الوصف الترتيبي لوحدات الجملة، والعمود الثاني يعرض وحدات الجملة على حدة ثم في حالة تركيبها مع وحدات أخرى، والعمود الثالث نوع المركبات، والعمود الرابع الوصف التصاعدي لعقد المركبات الذي يبدأ من العدد ٥٠٠.

٣-١-١-٢-١ خصائص هيكلة نيجرا في التمثيل النحوي

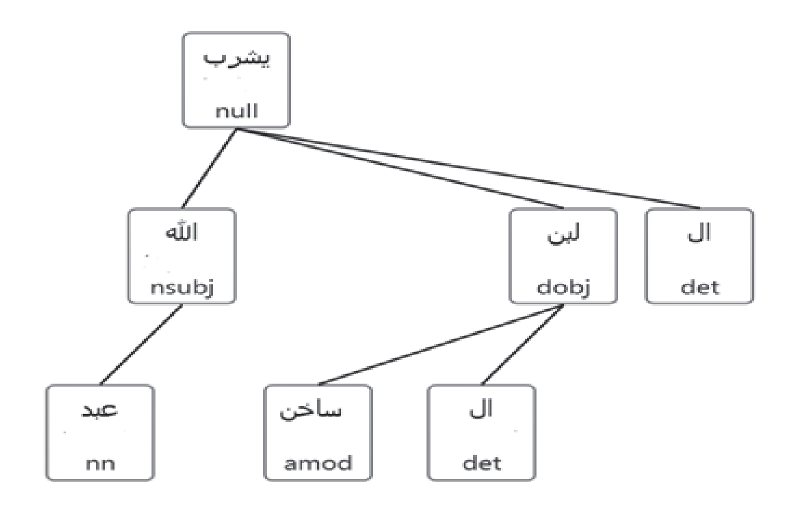
- سهولة قراءة البنية الشجرية، ومن ثم إمكانية التعديل أو الحذف أو الاستبدال داخل قاعدة البيانات.
- إمكانية استخلاص العلاقات التركيبية بسهولة دون استخدام أي أدوات حاسوبية مساعدة، حيث يتم تخزينها باستخدام أدوات قواعد البيانات مثل مايكروسوفت أكسس Microsoft Access أو إس كيو إل الاستعلامية SQL database.
- إمكانية توصيفها بلغة التوصيف المثالية القابلة للامتداد XML، إذ لا تتطلب ملفًّا تعريفيًّا (DTD) Document Type Definitions يحتوي على عناصر الوثيقة.
- أكثر ملاءمة من نظام التقويس في التعامل مع تقنيات التعلّم الآلي Machine Learning.
- إمكانية تحويلها إلى صورة – مرئية – شجرية مرة أخرى.

٣-١-٢ التمثيل الاعتمادي Dependency Representation

هو بنية شجرية منظمة، تنتظم فيها كلمات الجملة في شكل علاقات مكونة من مسيطر وتابع أو عدة توابع، بحيث تظهر فيها الكلمات كعُقَد نهائية Terminal فقط.(١) وهو ما يعرف بأشجار بنية الاعتمادية Dependency Structure Trees.

وتتشكل هذه البنية من خطوط متفرعة متجهة إلى الأسفل، تحكم تشعبها العلاقات القائمة في الجملة بين وحداتها، كما يتوقف تدرَّج أسهم تلك الخطوط بين وحدات الجملة على نوع العلاقات الحاصلة، إذ يتطابق التدرج مع الكلمات من حيث كونها مباشرة (أي مرتبطة بنقطة التمركز الرئيسي) أو غير مباشرة.

ويوضح الشكل رقم ٢٩ التدرج في البنية الشجرية الاعتمادية من خلال وحدات جملة "يشرب عبد الله اللبن الساخن".



الشكل: ٢٩ تمثيل العنونة الاعتمادية(٢) اعتمادًا على فئات نوح سميث.(٣)

---

(1) Habash, N & Farag, R & Roth, R. (2009). Syntactic Annotation in Columbia Arabic Treebank, In proceedings of the 2nd International Conference on Arabic Language Resources Tools MEDAR).P.129.

(٢) باستخدام أداة Dependency Viewer مفتوحة المصدر للباحث الصيني جوانجشاو Guangchao Tang بجامعة نانجينغ Nanjing university الصينية.

(٣) حيث تعني Null= لا علاقة، وNN= اسم (جزء أولي من مركب نحوي)، وNsubj= علاقة الفاعلية (الجزء المكمل للمركب)، وdet= أداة، وdobj= علاقة المفعولية، وamod= علاقة الوصفية.

تتدرّج الوحدات من نقطة التمركز الرئيسي ألا وهي الفعل تبعًا لمدى ارتباطها بتلك النقطة، فنجد في المثال السابق تدرج وحدتي " عبد الله" و " اللبن" في مرتبة مباشرة للفعل، حيث يحتلان موقعي الفاعلية والمفعولية، أما وحدة " الساخن" فهي في مرتبة مباشرة لوحدة " اللبن"، حيث إنها صفة لمفعولٍ.

ويتم التمثيل الاعتمادي في إطار نموذج تحليلي؛ ليحدد التعلق Attachment الذي يعنى بتوجيه العلاقات بين الرأس والتابع في الجملة، وبالمحتوى النحوي؛ ليحدد نوع تلك العلاقات labels (الفاعلية، المفعولية، الوصف، الإضافة...).[١]

ويتم نظم هذه البنية الاعتمادية في صفوف وأعمدة، ولعل أشهر طريقة في نظمها هي:

٣-١-٢-١ هيكلة كُنل CoNLL Format

هي إحدى الطرق الشائعة في البناء الشجري ذي التمثيل الاعتمادي، تقوم على توصيف البنية الشجرية الاعتمادية وتنظيمها في صفوف وأعمدة، إلا أن الأخيرة ترتبط بعدد محدد يحدده الهدف المنشود الذي يقرره المؤتمر السنوي لتعليم اللغة الطبيعية The Conference on Natural Language Learning (CoNLL) للآليات من خلال تقنيات التعلم الآلي والإحصائي، حيث يهدف هذا المؤتمر إلى تنظيم مهمة مشتركة Shared-Task لتقييم أنظمة معالجة اللغات الطبيعية ولا سيما المحللات النحوية Dependency Parsers، فيقوم بإعطاء المشاركين مدونة لغوية معنونة، متسقة من حيث عدد الأعمدة التي ترتبط بحجم المعلومات اللغوية المقدمة في المدونة؛ ليقسموها إلى جزءين، جزء للتدريب Training set، يتم به تدريب نماذجهم الإحصائية، وجزء لتقييم هذه النماذج بعد تدريبها Testing set.[٢] وقد أدَّى ذلك إلى ارتفاع مؤشرات تقييم المحللات النحوية المقدمة لأكثر من ثلاث عشرة لغة.[٣]

وسار العديد من صناع عنونة المدونات النحوية ذي التمثيل الاعتمادي على نهج

(1) Habash, N & Farag, R & Roth, R. (2009). Ibid. P.126.

(2) Nugues, P.M. (2014). Language Processing with perl and prolog. second Edition. springer.P.199.

(3) Haji˘c, J & Ciaramita, M & Johansson, R. et al,. (2009). The CoNLL-2009 Shared Task: Syntactic and Semantic Dependencies in Multiple Languages In Proceedings of the 13th CoNLL-2009, June 4-5 , Boulder, Colorado.

هيكلة كنل؛ للمشاركة بمدوناتهم النحوية في هذا المؤتمر العالمي، فضلا عن أنها أشبه ما تكون بصيغة عالمية موحّدة.

وقد تختلف صيغة تنسيق البيانات – من حيث عدد الأعمدة التي تنتهجها المدونات المعنونة – من عام لآخر حسب الهدف الذي ينشده المؤتمر كل عام، إلا أن هناك صيغة شائعة تشتمل على عشرة أعمدة ربما كانت كافية لاحتواء جل المدونات النحوية ذي التمثيل العلاقي باختلاف أهدافها.

ويوضح الجدول رقم ٣٠ هيكلة كنل لوحدات جملة "وقال رود لارسن متحدثا للإذاعة العامة النرويجية أنه يمكن الآن تشبيه المنطقة من عدة نواح بالبارود المشتعل". المقتبسة من بنك كولومبيا الاعتمادي CATiB.[١]

F5	F4	Functions	Relations	F3	F2	POS	F1	Word	ID
_	_	MOD	2	_	_	PRT	_	و+	1
_	_	---	0	_	_	VRB	_	قال	2
_	_	SBJ	2	_	_	PROP	_	رود	3
_	_	---	3	_	_	PROP	_	لارسن	4
_	_	MOD	2	_	_	NOM	_	متحدثا	5
_	_	MOD	5	_	_	PRT	_	ل+	6
_	_	OBJ	6	_	_	PROP	_	الإذاعة	7
_	_	MOD	7	_	_	PROP	_	العامة	8
_	_	MOD	7	_	_	PROP	_	النروجية	9
_	_	OBJ	2	_	_	PRT	_	أن	10
_	_	SBJ	10	_	_	NOM	_	هـ+	11
_	_	PRD	10	_	_	VRB-PASS	_	يمكن	12
_	_	MOD	12	_	_	NOM	_	الآن	13
_	_	SBJ	12	_	_	NOM	_	تشبيه	14

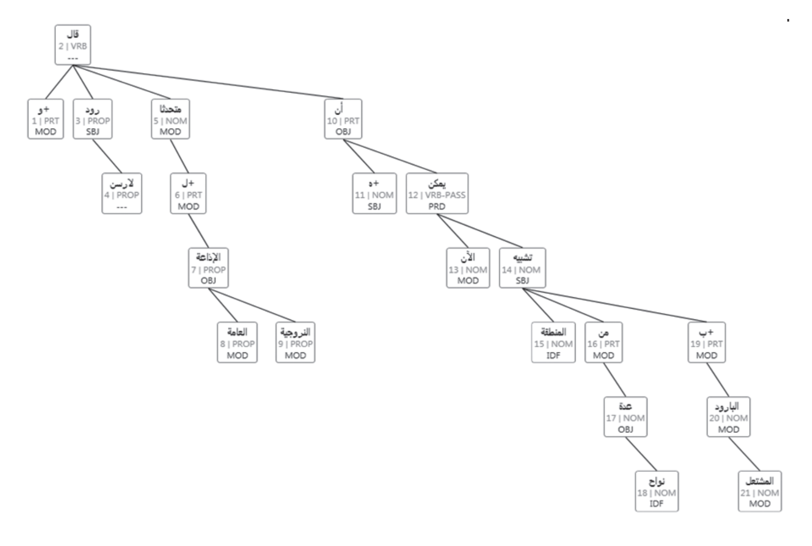
(١) قد طلبت من الأستاذ الدكتور نزار حبش – أستاذ معالجة اللغات الطبيعية بجامعة نيويورك أبو ظبي – أن يرسل لي عينة من مدونتهم الموصفة نحويًا للاطلاع عليها، فأرسل لي هذا النموذج ، له جزيل الشكر.

F5	F4	Functions	Relations	F3	F2	POS	F1	Word	ID
_	_	IDF	14	_	_	NOM	_	المنطقة	15
_	_	MOD	14	_	_	PRT	_	من	16
_	_	OBJ	16	_	_	NOM	_	عدة	17
_	_	IDF	17	_	_	NOM	_	نواح	18
_	_	MOD	14	_	_	PRT	_	ب+	19
_	_	MOD	19	_	_	NOM	_	البارود	20
_	_	MOD	20	_	_	NOM	_	المشتعل	21

الجدول: ٣٠ من عنونة بنك كولومبيا الشجري

فالعمود الأول يُعنى بالوصف الترتيبي لوحدات الجملة، أما الثاني فيعرض وحدات الجملة نفسها في ترتيب تنازلي، والرابع يعرض الأقسام الكلامية المتبعة في التحليل، والسابع يعرض العلاقات بين التابع والمتبوع في صورة رقمية، والثامن يعرض نوع العلاقة لكل وحدة في الجملة، بينما الأعمدة الخمسة (الثالث والخامس والسادس والتاسع والعاشر) خالية من السمات Features، إذ يتناول بنك كولومبيا خمس أعمدة فحسب.

وبتحويلها للشكل الشجري:



شكل ٣٠: تمثيل البنية الشجرية لمثال من بنك كولومبيا الشجري

٣-١-٣ التمثيل النحوي بين المكونية والاعتمادية

لقد أصبح التمثيل المكوني والاعتمادي من وجهة نظر حاسوبية وجهين لعملة واحدة، فكلاهما يعرضان النظام الكلي للغة لكن بطرق مختلفة، بل يتحدد معيار التفاضل بينهما عند تلك الوجهة في مقدرة أحدهما على توفير المعطيات اللغوية اللازمة للفهم الأتوماتي للنصوص اللغوية، أما من وجهة النظر اللغوية فكانت الفلسفة اللغوية ملاذًا للتمييز والتنظير لكليهما. وثمة فروق بين التمثيلين يوضحها الجدول رقم ٣١:

التمثيل المكوني	التمثيل الاعتمادي
يعتمد في بنائه على عقود المركبات أو العبارات.	يعتمد في بنائه على الكلمات والعلاقة بينها وبين الكلمة الرأس.
تظهر فيه الكلمات كعقد نهائية والمركبات كعقد غير نهائية.	تظهر فيه الكلمات كعقد نهائية فقط.
تظهر الوظائف النحوية والدلالية – غالبا – على عقود المركبات.	تظهر الوظائف النحوية والدلالية على عقود الكلمات.
يظهر فيه – غالبا – الفئات الفارغة، وكذلك الإشارة المرجعية للضمير.	لا يظهر فيه – غالبا – الفئات الفارغة، وكذلك الإشارة المرجعية للضمير.
يتم نظمه من خلال الأقواس الرياضية.	يتم نظمه من خلال الصفوف والأعمدة.
تعبر عن العلاقات بين أجزاء الجملة بالمركبات أو العبارات (المركب الاسمي، المركب الفعلي،...)	تعبر عن العلاقات بين أجزاء الجملة المتمثلة في الأسهم بالأرقام الاعتمادية.

الجدول: ٣١ فروق بين التمثيل المكوني والتمثيل الاعتمادي

٣-١-٣-١ تحويل التمثيل المكوني إلى الاعتمادي والعكس

يقصد بالتحويل نقل توصيفات Annotations بنية المكونات إلى بنية الاعتماديات أو العكس آليًّا، بحيث يتم التعبير عن المحتوى نفسه في التمثيل المحول إليه.[1] ولما شهد

(1) Rambow, O. (2010). The Simple Truth about Dependency and Phrase Structure Representations.Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the ACL, Los Angeles, California.

الدرس النحوي في بداية العقد الأول من القرن الحالي تطوّرًا كبيرًا في التحليل الاعتمادي Dependency Parsing ؛ فقد لجأ الباحثون إلى تحويل التمثيل المكوني إلى اعتمادي في الوقت الذي ارتكزت فيه التحليلات النحوية للعديد من اللغات على بنية المكونات.

ويقتضي التحويل التبصّر الجيد في المبادئ التوجيهية Guidelines لتحليل بنية المصدر وتحليل بنية الهدف، ثم تحديد الهدف من كلا التمثيلين؛ لتلمّس مواضع التوازي Mapping بين المعطيات النحوية في التمثيلين التي يُنطلق منها في بناء الخوارزميات اللازمة للتحليل الآلي.

ثمة أسباب قد تعوق عملية التحويل من بنية المصدر إلى بنية الهدف، منها: [(١)]

- إذا كان المصدر يحتوي على معلومات نحوية أقل من الحد المطلوب في بنية الهدف، على سبيل المثال التحويل من بنية مكونات لا تميّز بين حرف الجر والأداة إلى بنية الاعتماديات، حتمًا ستؤدي إلى خلل في تعيين بعض العلاقات بين أجزاء الجملة.

- إذا كانت بنية المصدر غير متسقة أو غير محكومة بنظرية نحوية.

ويرى أون رامبو Owen Rambo أن التحويل من بنية المكونات إلى بنية الاعتماديات أكثر سهولة من بنية الاعتماديات إلى بنية المكونات، في حين التحويل من الاعتماديات إلى المكونات يكون أكثر إفادة لتقنيات التعلم الآلي.[(٢)]

وقد تعددت الأدوات الحاسوبية في هذا المضمار لتحويل بنية المكونات إلى بنية الاعتماديات، ومع ذلك لا تصلح لكافة التحويلات في هذا الإطار إلا بتغيير المتغيرات Parameters التي تخضع للهدف المنشود، ومن هذه الأدوات:

١- محول بنك بنسلفانيا Penn Converter [(٣)]

وقد طوَّرت وحدة تقنية اللغة Language Technology بجامعة لاند هذه الأداة؛ لتحويل البنية المكونية المستخدمة في نمط بنسلفانيا إلى أشجار الاعتمادية للغة الإنجليزية في صيغة تنسيقات كنل ٢٠٠٧، ٢٠٠٨، ٢٠٠٩. وقد كتبت بلغة جافا java البرمجية،

(1) Rambow, O. (2010). Ibid.

(2) Rambow, O. (2010). Ibid.

(٣) يمكن تحميل الأداة من خلال هذا الرابط:
treebank_converter. (2007). http://nlp.cs.lth.se/software/treebank_converter

مدعمة العديد من بيئات التشغيل MS Windows, UNIX Distributions .

٢- المنسق من المكونية إلى الاعتمادية Constituency to Dependency Mapper[١]

وقد طوّر الدكتور نزار حبش هذه الأداة أثناء بنائه لبنك كولومبيا الاعتمادي CATiB؛ وذلك لتحويل بنك بنسلفانيا المكوني ATB إلى النهج الاعتمادي. وقد كُتبتْ هذه الأداة بلغة بيرل Perl البرمجية، مدعمة العديد من بيئات التشغيل MS Windows , UNIX Distributions .

ويمكن توضيح عملية التحويل من خلال عرض المثال المرفق مع الأداة:

"اطلع النائب العام التمييزي عدنان عضوم على التحقيقات الجارية في حادث اختفاء المهندس في التنظيم المدني وديع أبي راشد قبل أربعة أيام. وأفادت مصادر مطلعة على التحقيق أنه يجري التوسع فيه مع ثلاثة عناصر من قوى الأمن الداخلي احيلوا على مفرزة التحري لاستيضاحهم طبيعة العلاقة التي تربطهم بالمختفي." جريدة النهار.

```
(S
  (VP (PV+PVSUFF_SUBJ:3MS ATlE)
    (NP-SBJ
      (NP (DET+NOUN+CASE_DEF_NOM AlnA}b) (DET+ADJ+CASE_DEF_NOM AlEAm)
(DET+ADJ+CASE_DEF_NOM Altmyyzy))
      (NP (NOUN_PROP EdnAn) (NOUN_PROP EDwm)))
    (PP-CLR (PREP ElY)
      (NP
        (NP (DET+NOUN+NSUFF_FEM_PL+CASE_DEF_GEN AltHqyqAt)
(DET+ADJ+NSUFF_FEM_SG+CASE_DEF_GEN AljAryp))
        (PP (PREP fy)
          (NP (NOUN+CASE_DEF_GEN HAdv)
            (NP
              (NP (NOUN+CASE_DEF_GEN AxtfA')
                (NP
                  (NP
                    (NP (DET+NOUN+CASE_DEF_GEN Almhnds))
                    (PP (PREP fy)
                      (NP (DET+NOUN+CASE_DEF_GEN AltnZym)
(DET+ADJ+CASE_DEF_GEN Almdny))))
                  (NP (NOUN_PROP wdyE) (NOUN_PROP Aby) (NOUN_PROP
rA$d))))
              (NP-TMP (NOUN+CASE_DEF_ACC qbl)
                (NP (NOUN_NUM+NSUFF_FEM_SG+CASE_DEF_GEN ArbEp)
                  (NP (NOUN+CASE_INDEF_GEN AyAm))))))))))
  (PUNC .))
```

شكل:٣١ نموذج من بنك بنسلفانيا الشجري[٢]

(١) يمكن تحميل الأداة من خلال هذا الرابط:
C2D-v0.7. (2009). http://www1.ccls.columbia.edu/~habash/C2D-v0.7.tar.gz

(٢) تنتهج مؤسسة شبكة البيانات اللغوية نظام باكولتر في ترميز الحروف العربية.

وتظهر نتيجة التحويل من التمثيل المكوني إلى التمثيل الاعتمادي بهذا الجدول:

```
1      ATlE  VRB   0      ---    [WORD:ATlE,LEXEME:,VRB/
VBD/PV+PVSUFF_SUBJ:3MS]
2      AlnA}b      NOM    1      SBJ    [WORD:AlnA}
b,LEXEME:,DASHTAG:SBJ,NOM/DT+NN/DET+NOUN+CASE_DEF_NOM]
3      AlEAm NOM   2      MOD    [WORD:AlEAm,LEXEME:,NOM/
DT+JJ/DET+ADJ+CASE_DEF_NOM]
4      Altmyyzy    NOM    2      MOD
[WORD:Altmyyzy,LEXEME:,NOM/DT+JJ/DET+ADJ+CASE_DEF_NOM]
5      EdnAn PROP  2      MOD    [WORD:EdnAn,LEXEME:,PROP/
NNP/NOUN_PROP]
6      EDwm  PROP  5      ---    [WORD:EDwm,LEXEME:,PROP/
NNP/NOUN_PROP]
7      ElY   PRT   1      MOD    [WORD:ElY,LEXEME:,DASHTAG:
CLR,PRT/IN/PREP]
8      AltHqyqAt   NOM    7      OBJ
[WORD:AltHqyqAt,LEXEME:,NOM/DT+NNS/DET+NOUN+NSUFF_FEM_
PL+CASE_DEF_GEN]
9      AljAryp     NOM    8      MOD
[WORD:AljAryp,LEXEME:,NOM/DT+JJ/DET+ADJ+NSUFF_FEM_
SG+CASE_DEF_GEN]
10     fy    PRT   8      MOD    [WORD:fy,LEXEME:,PRT/IN/
PREP]
11     HAdv  NOM   10     OBJ    [WORD:HAdv,LEXEME:,NOM/NN/
NOUN+CASE_DEF_GEN]
12     AxtfA'      NOM    11     IDF
[WORD:AxtfA',LEXEME:,NOM/NN/NOUN+CASE_DEF_GEN]
13     Almhnds     NOM    12     IDF
[WORD:Almhnds,LEXEME:,NOM/DT+NN/DET+NOUN+CASE_DEF_GEN]
14     fy    PRT   13     MOD    [WORD:fy,LEXEME:,PRT/IN/
PREP]
15     AltnZym     NOM    14     OBJ
[WORD:AltnZym,LEXEME:,NOM/DT+NN/DET+NOUN+CASE_DEF_GEN]
16     Almdny      NOM    15     MOD
[WORD:Almdny,LEXEME:,NOM/DT+JJ/DET+ADJ+CASE_DEF_GEN]
17     wdyE  PROP  13     MOD    [WORD:wdyE,LEXEME:,PROP/
NNP/NOUN_PROP]
18     Aby   PROP  17     ---    [WORD:Aby,LEXEME:,PROP/
NNP/NOUN_PROP]
19     rA$d  PROP  18     ---    [WORD:rA$d,LEXEME:,PROP/
NNP/NOUN_PROP]
20     qbl   NOM   12     MOD    [WORD:qbl,LEXEME:,DASHTAG:
TMP,NOM/NN/NOUN+CASE_DEF_ACC]
21     ArbEp NOM   20     IDF    [WORD:ArbEp,LEXEME:,NOM/
NN/NOUN_NUM+NSUFF_FEM_SG+CASE_DEF_GEN]
22     AyAm  NOM   21     IDF    [WORD:AyAm,LEXEME:,NOM/NN/
NOUN+CASE_INDEF_GEN]
23     .     PNX   1      MOD    [WORD:.,LEXEME:,PNX/PUNC/
PUNC]
```

الجدول: ٣٢ التمثيل الاعتمادي المحول إليه من التمثيل المكوني

٣-٢ المحتوى النحوي Syntactic Content

يشير مصطلح المحتوى النحوي إلى المعطيات النحوية التي أنتجتها الميتا – لغة Meta Language من خلال التجريد والتوصيف للظاهرة اللغوية؛ لتشمل المقولات والمكونات والرموز التي تعبّر عن القواعد والعلاقات النحوية، والتي يتعذر بدونها تمثيل النظرية النحوية، كما لا يمكن – أيضًا – أن تمثَّل وحدها دون نظرية نحوية، لكن يمكن أن نشير إليها فحسب أثناء العملية التعليمية. [(١)]

وقد كان لعلمائنا اللغويين والنحويين القدامى فضل السبق في وضع منظومة رمزية لوصف القواعد والمعطيات النحوية، وقد بثوها في متونهم وكتبهم النحوية، ومع ذلك لم نجار هذه الأسبيقة إزاء الانفجار التنظيري، بل التزمنا الصمت تجاه التنظير اللغوي الحديث، ويعبر الدكتور نبيل على عن أزمة تنظيرنا النحوي الآن بقوله" ولا نذيع سرا بقولنا إن لغة وصف اللغة العربية قد تهرأت وتخلفت حتى على مستوى المفاهيم الأساسية، مثل مصطلح النحو والتركيب". [(٢)]

"يلجأ النحو العام بحكم طبيعته الشاملة إلى التجريد، ويميل إلى استخدام عدد محدود من الرموز والاصطلاحات، حيث تزداد قوة التنظير، ووجاهته، كلما أصبحت " لغة وصف اللغة" أكثر تقيدا وتجريدًا، فكلما توسعنا في استخدام الرموز والاصطلاحات، وتمادينا في تفريعاتنا وتصنيفاتنا، تاه في تفاصيلها لب الظاهرة النحوية، وتقلصت قدرتنا على وضع اليد أو القبض على عمومياتها وكلياتها." [(٣)]

وتتقيد المدونة اللغوية المعنية بالتحليل أو بالوصف للعلاقات التركيبية بعدد محدد من المعطيات النحوية Tag set التي تخضع للهدف المنشود من البناء، كما تتقيد المعطيات بنوع التمثيل المتوخي في التحليل، إذ يقتضي التمثيل المكوني معطياتٍ حول الوحدات المكونة للجملة التي لا يتطلبها التمثيل الاعتمادي، ويقتضي التمثيل الاعتمادي بيان العلاقات النحوية الممثلة التي قد لا يقتضيه التمثيل المكوني.

(1) Rambow, O. (2010). Ibid.

(٢) علي، (نبيل)، حجازي، (نادية): الفجوة الرقمية، ص ٣٦٤.

(٣) علي، (نبيل): اللغة العربية والحاسوب، ص٣٣٦، ٣٣٧.

وتتمحور جل المعطيات النحوية التي تم بها توصيف العديد من اللغات، حول ما يلي:

- السمات التركيبية Syntactic Tagset وتشتمل على علامات لأنواع الجمل، وأنواع المركبات أو الوحدات المكونة للجملة.
- السمات الوظيفية Functional Tagset وتشتمل على علامات للوظائف النحوية والدلالية والتداولية للوحدات أو المكونات، وكذلك خواص أجزاء النصوص.
- الإشارات المشتركة Co-Indexing لتمييز المركبات أو الوحدات الاسمية بعلامات محددة في الجملة.
- الفصائل الفارغة Empty categories وتشمل علامات للضمائر أو الأسماء أو المكونات المحذوفة في الجملة.

٣-٣ النظرية النحوية The syntactic theory

النظرية هي "مجموعة متجانسة من المبادئ والأسس والقواعد، التي ينتظمها مفهوم مستوعب للكلام وأحكامه، قادر على إيصال الدلالات الصحيحة للمخاطب"[(١)] وتنطلق هذه المبادئ من تمثيل المعنى فى إطار المعطيات النحوية التي تعدّ خصائص عامة في كثير من اللغات أو في اللغة المعينة.

وتجدر الإشارة إلى أن اختيار نوع التمثيل النحوي يحدد اتجاهات النظرية المتوخاة ما بين النظام التحليلي أو التوليدي، حيث يتلاءم النظام التحليلي مع التمثيل الاعتمادي، بينما يتلاءم النظام التوليدي مع التمثيل المكوني.

وتدور النظريات النحوية في فلكين أساسين، فَلَك يعنى بمبادئ الوصف النحوي للغةٍ مخصوصةٍ، وفلك آخر يعنى بالمبادئ التفسيرية للظاهرة النحوية في جميع اللغات الإنسانية. أو بقول آخر، تدور النظريات النحوية فيما أرساه دي سوسير بجعل اللغة ظاهرة زمكانية يمكن وصفها وصفًا آنيًّا، وفيما افترضه تشومسكي حول النحو الكلي UG بأنه موجود في أذهان الأطفال منذ الولادة، ويتمثل في مجموعة من المبادئ والمحددات.

ومن ثم تبلورت النظريات النحوية - في إطار التمثيل النحوي - في نوعين:

---

(١) عفيفي(أحمد مصطفى):النظرية النحوية-المفاهيم والتحديات، وقائع مؤتمر«العربية وقرن من الدرس النحوي» دار العلوم-القاهرة» ٢٠٠٣م، ص ١٩٢.

٣-٣-١ النظرية الوصفية: هي ضرب من بيان مبادئ طرق انعقاد الربط والارتباط بصورة محكمة في الجملة، وكذلك بيان موقع المكونات أو الوحدات في الجملة. ويتم توظيفها في التمثيلين النحويين من خلال:

٣-٣-١-١ البنية المكونية Constituency Structure

تنطلق النظرية الوصفية في ضوء التمثيل المكوني من النحو المتحرر من السياق[١] الذي حدده تشومسكي لتوصيف الظواهر اللغوية.

٣-٣-١-١-١ النحو المتحرر من السياق Context Free Grammar

هو النظام الرياضي الأكثر شهرةً لنمذجة بنية الوحدات أو المكونات للغة الإنجليزية، وللغات الطبيعية الأخرى، وهو ما يعرف بـ نحو بنية العبارة Phrase-Structure Grammar الذي اقترحه تشومسكي في كتابه التراكيب النحوية.[٢]

وتشبه هذه الصورية صيغة باكوس نور Backus-Naur أو BNF من حيث قواعد إعادة الكتابة (Rewrite_Rules)، التي يفترض في صياغتها عدم اعتماد أي عنصر فيها على ما يسبقه أو يلحقه من عناصر، أو بقول آخر، ليس له حساسية للسياق النحوي الذي يرد به.[٣]

ويتم صياغة الوحدات أو المكونات النحوية في ضوء عدد من قواعد بنية العبارة التي تشترك مع الفئات النحوية؛ لتكوين هذه الوحدات أو المركبات المعبرة عن أركان الجملة، ومن هذه الفئات الأساسية:

- الاسم Noun
- الفعل Verb
- الصفة Adjective
- حال Adverb
- حرف جر Preposition

(١) وتفرعت عدة شكليات formalisms من النحو المتحرر من السياق لوصف بنية الجملة منها: نحو الأشجار المتجاورة (TAG) Grammars Tree Adjoining، نحو بنية المقولات النحوية المعتمد على الرأس Head driven Phrase-Structure Grammar (HPSG)، والنحو الوظيفي المعجمي Lexical Functional Grammar (LFG)، والنحو المقولي Categorical Grammar (CG).

(2) Jurafsky, D. & Martin, J. (2007). Ibid. P.433.

(٣) علي(نبيل): السابق، ص ٣٤٢.

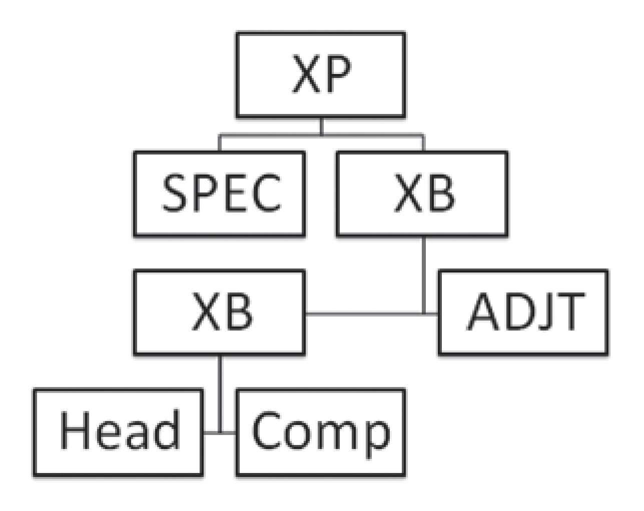
- أداة تحديد Determiner
- رابط Conjunction

ويتم تكوين القواعد من شقين يتخللهما سهمٌ، يمكن أن نعبّر عنهما بهذه الصيغة الرياضية W ←V، حيث V تعني اسم المركب، أما W فتعنى الأجزاء المكونة لهذا المركب التي قد تكون عبارة عن مركبات أو مزيج من مركبات وفئات نحوية، أو فئات نحوية فقط.[1] ومن أمثلة قواعد صياغة المركبات:

المركب الاسمي ← اسم + مركب اسمي	NP →N+ NP
مركب الجر ← حرف جر + مركب اسمي	PP →P+ NP
المركب الفعلي ← فعل + مركب اسمي+ (مركب جر)	VP →V+NP (PP)

يلاحظ في صياغة هذه المركبات أن لكل مركب رأسًا Head يميزه، ومكملًا Complement يتممه، على سبيل المثال، فالمركب الاسمي رأسه الاسم، والمركب الفعلي رأسه الفعل، ومركب الجر رأسه حرف الجر، بينما المكمل هو الأخت لرأس المركب في التمثيل الشجري، ويوضح الشكل رقم ٣٢ القوانين الشكلية لتنظيم قواعد بنية العبارة أو قواعد النحو المتحرر من السياق:

(1) Jurafsky, D. & Martin, J. (2007). Ibid. P.434.



شكل: ٣٢ المخطط التنظيمي لقواعد بنية العبارة

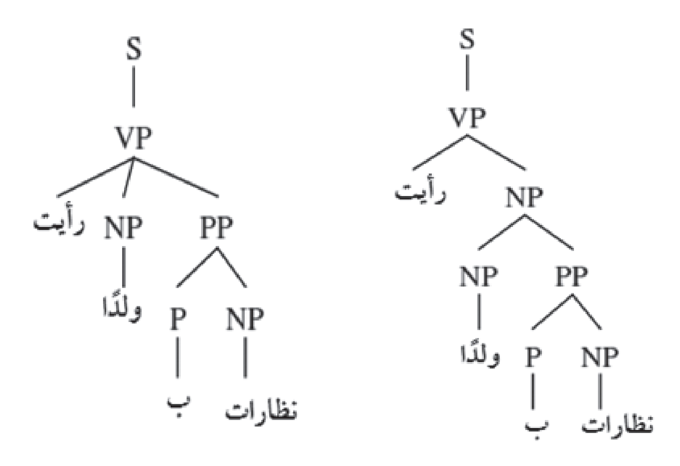
يجمع هذا الشكل القوالب التنظيمية لصياغة الجملة، حيث يحدد المحدد Specifier نوع المركب الذي يوازيه في التنظيم، "وهو في الغالب يقدم معلومات أكثر عن الرأس أو أن يجعل الرأس أكثر تخصيصًا"[١]، بينما الوصف Adjunct هو نوع من المركبات يصف رأس المركب، والمكمل COMP هو الأقرب للرأس. وهذه القوالب كانت المحكّ الأساسي في ظهور نظرية إكس بار X-bar أو السين البارية، إذ عمل تشومسكي على تقليص قوانين بنية العبارة وتحديدها في ضوء تلك القوالب أو ما تسمى بالمحددات Parameters.

لا يتم هذا التنظيم في صياغته للجملة إلا في ضوء المبادئ أو الأحكام التركيبية المقيدة ببنية العوامل Predicate-argument Structure التي تعادل في جوهرها نظرية العامل النحوي.[٢]

ويتضح تنظيم قواعد النحو المتحرر من السياق في ضوء الأحكام التركيبية المقيدة ببنية العوامل في المثال التالي الذي آثرت فيه أن يكون له أكثر من صورة؛ ليتضح التمييز بين التمثيلات المختلفة.

(١) بوول (جيفري): النظرية النحوية، ترجمة د. مرتضى باقر، المنظمة العربية للترجمة، ط١، ٢٠٠٩م، ص ١٢٤.

(2) Babby, L. (2009). The Syntax of Argument Structure. Cambridge University Press.P.11.



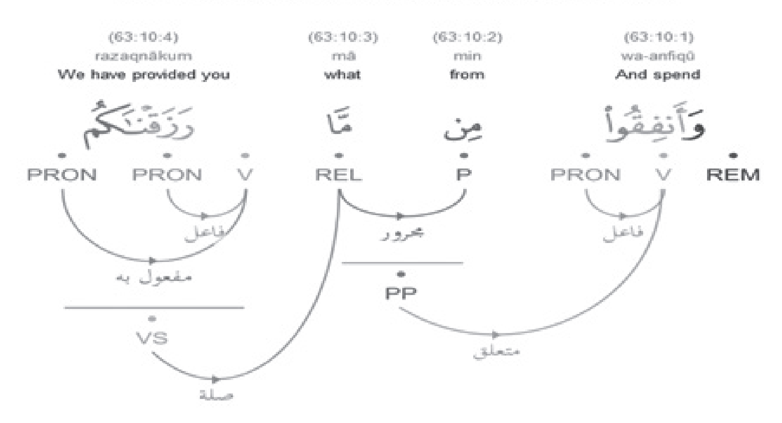
الشكل:٣٣ مثال على تنظيم الأحكام التركيبية المبنية على المعنى في إطار النحو المتحرر من السياق

يلاحظ في التمثيلين النحويين الأحكامَ التركيبية المبنية على المعنى، ففي الشكل الأول نجد أن مركب الجر PP هو وصف لرأس المركب الاسمي NP "ولدًا"، وهذا يعني أن النظارات صفة ملازمة للولد، أما الشكل الثاني فنجد أن مركب الجر PP هو مكمل لرأس المركب الفعلي VP " رأيت"[(١)]، وهذا يعني أن النظارات ملازمة للرّائي.

٣-٣-١-٢ البنية الاعتمادية Dependency Structure

تنطلق النظرية الوصفية في ضوء التمثيل الاعتمادي من النحو التقليدي القائم على الأحكام التركيبية المقيدة بفكرة العامل والمعمول أو الموضوع والمحمول، ولعل أصدق مثال على ذلك ما صنعه فريق معالجة اللغات الطبيعية بجامعة ليدز من بناء مدونة للنص القرآني معنونة بالعلاقات التركيبية ذي التمثيل الاعتمادي.

(١) يلاحظ عدم عنونة رأس المركب الفعلي برمز V، إذ لا يوجد احتمال آخر لنوع هذا الرأس غير الفعل أو شبه الفعل أو ما يقوم مقامه بعكس المركبات الأخرى، كما أن توصيف الفعل يحتل نصيبًا في عنونة الأقسام الكلامية، فلا داعي للتكرار هنا.



الشكل: ٣٤ نموذج بنية الاعتمادية «للنص القرآني»[(١)]

٣-٣-٢ النظرية التفسيرية:

«هي مجموعة المبادئ المنظمة التي ينبغي أن يلحظها البحث اللساني من حيث هي مشتركة بين اللغات وتلتزم بها اللغات».[(٢)] وهي ما تسمى بالقواعد الكلِّيَّة أو النحو الكلي الذي يقوم على المبادئ العامة General Principles المشتركة بين أنحاء اللغات. ويتم توظيفها – أيضًا – في التمثيلين النحويين من خلال:

٣-٣-٢-١ البنية المكونية Constituency Structure

تنطلق النظرية التفسيرية في ضوء التمثيل المكوني من نظرية السين البارية التي اقترحها تشومسكي[(٣)] بعد مرور خمسة أعوام من مؤلفه الأول بهدف إعادة صياغة البنية التكوينية لقواعد بنية العبارة؛ لتحديد التعميمات التي تحكم الفئات الداخلية لجميع اللغات الإنسانية.

---

(١) يمكن الاطلاع على المدونة القرآنية الموصفة من خلال هذا الرابط:
Quran Syntax. (2007). http://corpus.quran.com/treebank.jsp

(٢) زكريا، (ميشال): الألسنية التوليدية وقواعد اللغة العربية (النظرية الألسنية)، ط٢، المؤسسة الجامعية للدراسات والنشر والتوزيع، بيروت،١٩٨٦م، ص ٧٧.

(٣) قد أعاد كل من بريزنان Bresnan ١٩٧٦م، وإيمندز Emonds ١٩٧٦م، وجاكندوف Jackendoff ١٩٧٧م صياغة هذه النظرية بطرق مختلفة. ينظر: زفارت (جوست): البنيات التركيبية والبنيات الدلالية، علاقة الشكل بالمعنى. ترجمة د.عبد الواحد خيري، ط١، دار الحوار للنشر والتوزيع، سوريا، ٢٠٠٨م، ص ٢٤-٢٥.

٣-٣-٢-١-١ نظرية السين البارية X-bar Theory

تفترض هذه النظرية وجود ملامح تركيبية مشتركة بين جميع اللغات الإنسانية بناءً على افتراضيات وجود نمط معين منشأ بالذهن البشري، وقد عبّر تشومسكي عن أنساق هذه الملامح التركيبية - من خلال هندسة التوازي بين الذهن والنحو - بتعيينه للمبادئ والمحددات اللغوية.

وتتمثل المبادئ اللغوية الحاكمة للمكون التركيبي في التمثيل الشجري في مبدأين أساسيين[١] :

- مبدأ الرأسية (Headedness Principle).

كل عقدة غير نهائية في البنية التركيبية تمثل إسقاطًا لرأس الكلمة.

- مبدأ الثنائية (Binarity Principle).

كل عقدة غير نهائية Non-Terminal في البنية التركيبية تكون ذات تشعب ثنائي. وبذلك يتم التخلص من الأبنية المسطحة Flat Structures التي تظهر في صياغة قوانين بنية العبارة التي أرساها تشومسكي في كتابه البنى التركيبية عام ١٩٦٥م.

وفي ضوء هذه المبادئ، تتحدد القوالب التنظيمية لصياغة قوانين بنية العبارة في إطار المحددات Parameters العامة[٢] التي تشمل جميع اللغات الإنسانية، ومن هذه القوالب أو القوانين العامة[٣]:

١ - قانون المحدد The Specifier Rule

يتكون المركب XP (حيث تعنى X أو س الفئة المقولية مثل الاسم، الفعل، الصفة ...) من محدد اختياري Spec (مثل أل التعريف، وحروف العطف، و ضمائر الوصل، ...) و س شرطة (X')، مع إمكانية تبديل المواقع ما بين المحدد وس شرطة، حيث تختلف اللغات فيما بينها في مواقع تواجد هذه الأجزاء.

---

(1) Radford, A. (2009). An Introduction to English Sentence Structure, University of Essex. P.43.

(٢) يقصد بالمحددات العامة هنا الإحاطة بقيود الرتبة التركيبية في اللغات الإنسانية؛ لذا سنجد في القوانين إمكانية تبديل المواقع بين المحدد وسين شرطة، وبين الوصف ورأس المركب، وبين المكمل والرأس؛ لتختار كل لغة المحددات الخاصة بها عند التحليل. على سبيل المثال مركبات الوصف Adjunct يمكن أن تأتي في اللغة الإنجليزية على يمين الرأس ويساره، وكذلك في العربية إذا كان مركب الوصف حالا، أما إذا كان صفة فتأتي بعد الرأس.

(3) Radford, A. (2009). Ibid. P.13.

XP →X' (Spec) أو XP →(Spec) X'

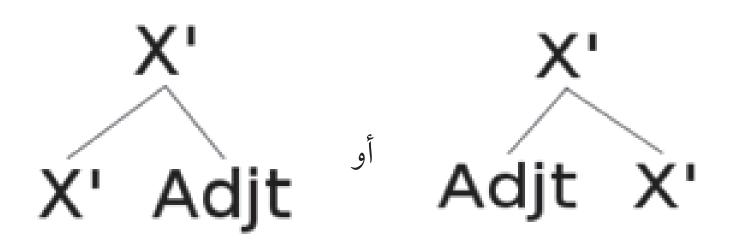


الشكل: ٣٥ التمثيل الشجري لقانون المحدد[1]

٢- قانون الوصف The Adjunct Rule

يأتي هذا القانون لإيضاح وصف رأس المركب، كما يتموقع هذا الوصف (مركب الصفة، مركب الحال، مركب الجر غير المكمل،...) على يمين أو يسار الرأس المتمثل في السين شرطة أو X'، وذلك لاختلاف مواقع الوصف في اللغات.

X' →X' Adjunct أو 'X' →Adjunct X



الشكل: ٣٦ التمثيل الشجري لقانون الوصف[2]

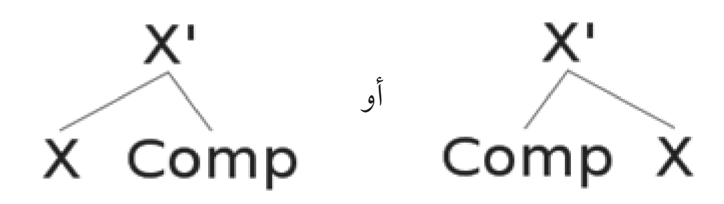
٣- قانون المكمل The Complement Rule

يأتي هذا القانون لإيضاح مكمل رأس المركب الذي يتمثل في المفعول به المباشر وغير المباشر، والمضاف إليه، وتمييز العدد،...، ويتموقع هذا المكمل – أيضًا – على يمين رأس المركب ويساره المتمثل في السين أو X.

(1) Koopman, H & Sportiche, D & Stabler, E. (2003). An Introduction to Syntactic Analysis and Theory . UCLA. P.125.

(2) Koopman, H. & Sportiche, D. & Stabler, E. (2003). Ibid. P.125.

X’ → Complement X   أو   X’ → X Complement



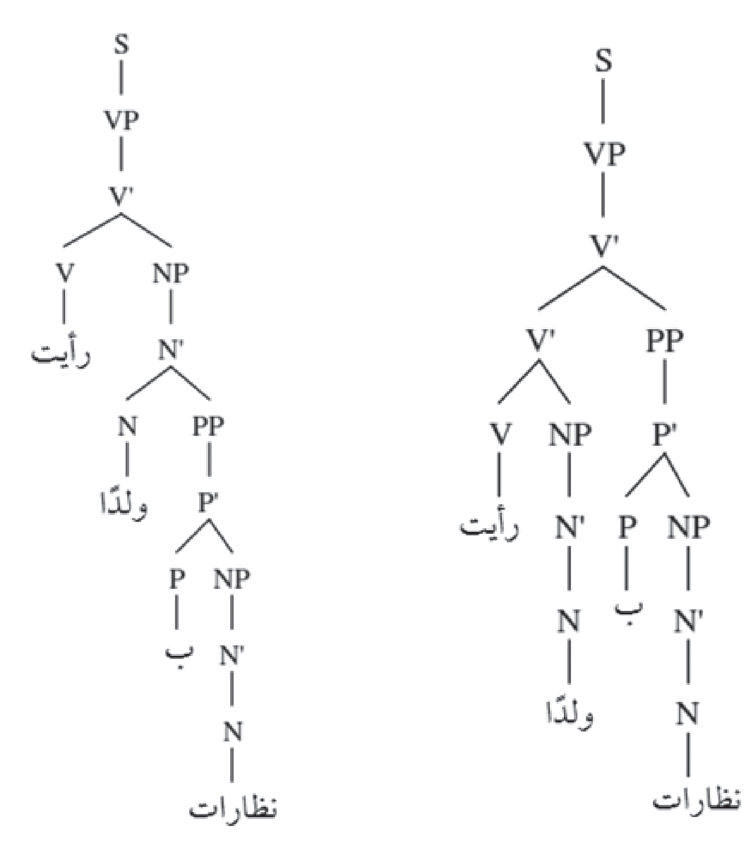
الشكل: ٣٧ التمثيل الشجري لقانون المكمل[١]

ويقتضي تنظيم هذه القوالب التنظيمية أو القوانين العامة الاعتمادَ على الأفكار العامة للغات الإنسانية: مثل الموضوع والمحمول، والمسند والمسند إليه وغيرها من الخصائص العالمية للغات.[٢] وهذا يؤكد مما لا يدع مجالا للشك أن هناك اتفاقا تامًّا بين النحو المتحرر للسياق ونظرية السين البارية في اعتمادهما على الموضوع والمحمول في تنظيم القوالب لصياغة الجملة، لكن الفارق الوحيد بينهما في طريقة صياغة تلك القوانين التوليدية، حيث كانت القوانين في النحو المتحرر من السياق كثيرة جدًّا، كما تختلف من لغة لأخري حسب رتبة الكلمات في الجملة، بينما في نظرية السين البارية محددة جدًّا، وتحتوي كل اللغات الإنسانية مع اختلاف بعض المحددات Parameters من لغة لأخرى.

وباختيار المحددات Parameters الخاصة للغة العربية، يمكننا أن نبيّن تنظيم القوالب - في ضوء الأحكام التركيبية المقيدة بالعامل والمعمول - في لغتنا العربية، وقد عمدت إلى اختيار المثال نفسه الذي أجريت عليه قواعد النحو المتحرر من السياق في ضوء الأحكام التركيبية؛ ليتضح الفارق بينهما في صورة تمثيلية.

(1) Koopman, H. & Sportiche, D. & Stabler, E. (2003). Ibid. P.125.

(٢) البنهساوي (حسام): نظرية النحو الكلي والتراكيب اللغوية العربية، مكتبة الثقافة الدينية، ٢٠٠٤م، ص ٦١-٦٢.



الشكل:٣٨ مثال على تنظيم الأحكام التركيبية المبنية على المعنى في إطار نظرية السين البارية

وبملاحظة الصورة الأولى للمثال نجد أن مركب الجر PP في موضع الوصف لرأس المركب الفعلي، حيث إنه بنت لـ V' وأخت لـ V' في التمثيل الشجري؛ ليدل على أن النظارات صفة ملازمة للولد، أما في الصورة الثانية للمثال نجد أن مركب الجر PP في موضع المكمل لرأس المركب، حيث إنه بنت لـ N' وأخت لـ N في التمثيل الشجري؛ ليدل على أن النظارات ملازمة للرّائي.

٣-٣-٣-٢ البنية الاعتمادية Dependency Structure

تنطلق النظرية التفسيرية في ضوء التمثيل الاعتمادي من النظرية الاعتمادية[1] التي

(١) تفرعت العديد من الشكليات formalisms من نظرية تنيير الاعتمادية، منها: نحو الكلمة Word Grammar (WG)، والوصف التوليدي الوظيفي Functional Generative Description (FGD)، ونظرية نص المعنى Meaning-Text Theory (MTT)، والنحو العلاقي Link Grammar (LG).

يعود بزوغها إلى اللغوي الفرنسي لوسيان تنيير(L.Tesnière) في كتابه عناصر النحو التركيبي"Eléments de Syntax Structurale, Paris 1959" الذي اتخذ صورته النهائية بعد دراسات سابقة في هذه المجال، ويعد هو المرتكز الفعلي لأسس نظرية النحو التركيبي لديه.[١]

٣-٣-٣-٢-١ نظرية الاعتمادية Dependency Theory

ترتكز هذه النظرية على مبادئ عامة تشترك فيها جميع اللغات الإنسانية؛ لإقامة نموذج نحويّ يساير كل الأنظمة اللغوية المختلفة من حيث انتظام عناصر الجملة في قالب محدد من العلاقات. وتتمثل هذه المبادئ بشكل عام في تحديد الرؤوس التركيبية باعتبارها حاكمة لمكملاتها أو توابعها، ويحتل الفعل قمة العناصر الحاكمة في الجملة.

وتتدرج الاعتمادات من قمة الفعل باعتباره المؤثر الأقوى في تأليف نواظم الكلم إلى العناصر الأساسية المباشرة له، ثم من تلك العناصر إلى عناصر تتبعها، وهكذا.أما إذا لم يكن الفعل موجودًا في بنية الجملة، فإن المركب الاسمي وحده قادر على أن يبني جملة يطلق عليها الجملة الاسمية من خلال مركزية الاسم في المركب.[٢]

وقد تبلورت العموميات اللغوية الاعتمادية بشكل تطبيقي على يد اللغوية ماري كاثرين دي مرنف Marie-Catherine de Marneffe أستاذة اللغويات بجامعة ولاية أهاويو Ohio State University بالولايات المتحدة الأمريكية، حيث مكنت من إقامة نموذج نحوي موحّد يتناول طرق تمثيل العلاقات النحوية في البنية الاعتمادية في جميع اللغات الإنسانية مع تقديم بعض المحددات Parameters التي تختلف فيها بعض اللغات، وكان ذلك إثر استقصائها للعلاقات النحوية في اللغة الإنجليزية؛ لاستخدامها آنذاك في اعتماديات ستانفورد التي كانت تهدف إلى تطبيقات فهم اللغة الطبيعية (NLU) Natural Language Understanding مثل المحللات النحوية، الترجمة الآلية، وإجابة الأسئلة، والمحاورة مع الحاسوب، وغيرها من التطبيقات.[٣]

(١) بحيري(سعيد حسن): نظرية التبعية في التحليل النحوي،مكتبة الأنجلو المصرية، ط١، ١٩٨٨م،ص.١٢.

(٢) السابق، ص ١٦٥.

(3) Marneffe, de & Dozat, T. & Sliverira, N. et al., (2014). Universal Stanford Dependencies: A cross-linguistic Typology. In Proceedings of 9th International Conference on Language Resources and Evaluation (LREC 2014).

وكان للنموذج النحوي الذي وضعته ماري مرنف أثر بالغ في اللسانيات الحاسوبية لا سيما التحليل النحوي Syntactic Parsing، فقد انطلق منه جوكايم نيفر، وريان ماكدونالد وغيرهم في بناء مشروع "الاعتمادات العالمية Universal Dependencies (UD) لعنونة اللغات الإنسانية بمنهجية اعتمادية موحدة؛ بهدف بناء محلل نحوي متعدد اللغات من أجل توحيد دقة النتائج التحليلية بين اللغات المختلفة؛ وذلك لتحسين نتائج الترجمة الآلية.(١)

وقد تضمن هذا المشروع إلى الآن ثمان وعشرين لغة(٢)، تتصدرهم اللغة الإنجليزية والألمانية والسويدية والصينية والعبرية، وتخلو منها اللغة العربية للأسف الشديد بسبب إغفالنا التام عن النظريات اللغوية الحديثة واللغويات الحاسوبية والمعرفية، وغيرها من مقتضيات عصرنا النهضوي، ولا نبالغ بقولنا إن قدرة لغتنا العربية على اللحاق بركب الثورة المعرفية، ومواجهة العولمة التي تمارس عليها ضغوطًا بسبب السيل المعلوماتي الجارف رهن بما يمكن إنجازه على صعيد معالجتها آليًّا.

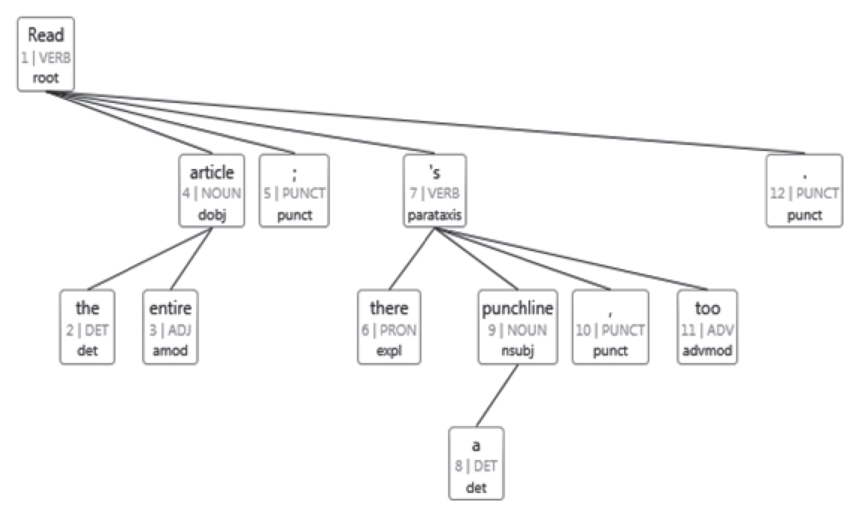
ولما كان هذا المشروع العالمي يخلو من لغتنا العربية، فقد تخيرت مثالا توضيحيًّا من مدونة ويب للغة الإنجليزية المعنونة بالاعتمادات العالمية(٣)؛ مستخدمًا أداة Dependency Viewer المفتوحة المصدر للباحث الصيني جوانجشاو Guangchao Tang لعرض هذا المثال في هيئة شجرية.

---

(1) McDonald, R. & Niver, J. et al., (2013). Universal Dependency Annotation For Multilingual Parsing. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics. Sofia, Bulgaria.

(٢) يمكن الاطلاع على موقع الاعتمادات العالمية من خلال هذا الرابط:
Universal Dependencies. (2014).
http://universaldependencies.github.io/docs/#language-u

(٣) للاطلاع على مدونة ويب العالمية للغة الإنجليزية من خلال هذا الرابط:
UD_English. (2014). https://github.com/UniversalDependencies/UD_English

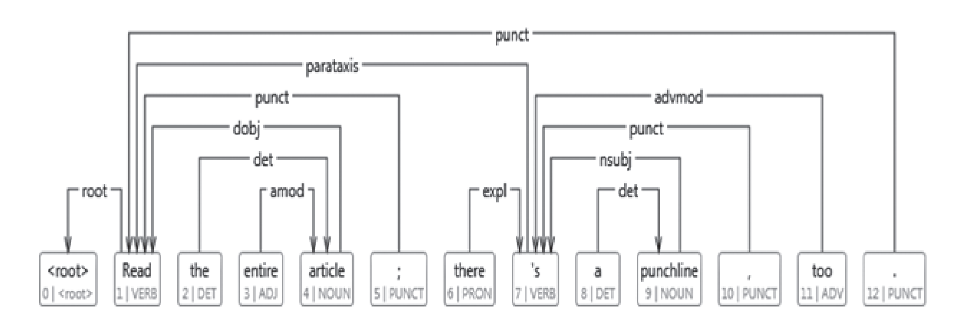


الشكل:٣٩ نموذج من مدونة ويب للغة الإنجليزية (في صورة شجرية)

.Read the entire article : there’s a punchline, too

اقرأ المقال كاملًا: ثمة روح فكاهية بنهاية النص.

أو في هيئة أفقية Horizontal:



الشكل:٤٠ نموذج من مدونة ويب للغة الإنجليزية (في صورة أفقية)

.Read the entire article : there’s a punchline, too

اقرأ المقالة كاملًا: ثمة روح فكاهية بنهاية النص.

٤ . عنونة المدونة اللغوية – عينة الدراسة – بالعلاقات التركيبية

لما كان الوصول إلى نظام تحليلي مكتمل الأركان، صارم الأحكام، من حيث عوامل التحليل اللغوي: التمثيل والنظرية والمحتوى، يعد أمرًا صعبًا نسبيًّا في وقتنا الحاليّ، لاعتماد تلك الأنظمة التحليلية – المتاحة – على المبادئ الافتراضية فحسب دون المساس بكل العلوم المعرفية Cognitive Science التي لها علاقة وثيقة باللغة، فقد كان لزامًا علينا أن نتوخى إحدى تلك الأنظمة في الكشف عن الأبنية التركيبية في النظام اللغوي موضوع الدراسة.

وتتمثل تلك الأنظمة التحليلية في القيام بدور المسوغ للقواعد التركيبية (النحوية) عن طريق تقديم معيار أو إجراء تقويمي، يمكن عن طريقه أن يتخير العالم اللغوي أفضل تلك الإجراءات لتكون صالحة دون سواها في تحليل المادة اللغوية، وليس يعني ذلك أن هذا الإجراء أو المعيار هو وحده الصحيح بصورة مطلقة، لكنه – بحق – أفضل الإجراءات في وصف اللغة وتحليلها.[١]

وتتحدد ملامح العنونة بالعلاقات التركيبية للمادة اللغوية عينة الدراسة في ضوء الأغراض المنشودة من البناء. وقد انطلقت الدراسة من التمثيل المكوني في تعيين العلاقات التركيبية لعدة أسباب:

- الأكثر تمثيلًا لشكل المعرفة اللغوية لاسيما القواعد النحوية في الذهن البشري (الفص الصدغي).[٢]
- بمثابة البنية التحتية التي تصلح لتمثيل المحتويين المكوني والاعتمادي.
- للحصول على سمات Features تركيبية ممزوجة من المكونية والاعتمادية؛ وذلك للإفادة منها في تدريب أساليب الذكاء الاصطناعي المختلفة.

ولما كان الهدف المنشود من البناء يتطلب توفير معلومات نحوية حول أجزاء التركيب الجملي في المدونة اللغوية، فقد اعتمدت الدراسة على النظرية الوصفية في تمثيل

---

(١) البهنساوي (حسام): القواعد التحويلية في ديوان حاتم الطائي، مكتبة الثقافة الدينية، القاهرة ، ١٩٩٢م، ص ٧٣.

(٢) قد استخدم مختبر اللسانيات العصبية الحاسوبية Computational Neurolinguistics Lab بجامعة ميشيغان الأمريكية University of Michigan هذا التمثيل المكوني باعتباره بديلا عن النموذج النحوي الذهني، وذلك لحساب عدد العُقد في التركيب الجملي مقارنة بعدد المشتبكات العصبية التي تتشكل داخل الذهن البشري أثناء نطق هذا التركيب الجملي. ينظر:
Hale, J & Callaway, F & Feldman, E. et al., (2014). Ibid.

العلاقات التركيبية باعتبارها ملاذًا لتوفير المعطيات اللغوية التي تمثل خصائص خاصة للغة المعينة، إذ يتنوع التوصيف حسبما يتنوع النظم التركيبي للجملة، بعكس ما تتيحه النظرية التفسيرية التي تعتمد على قوانين محددة، ومن ثم معلومات محددة عن الأجزاء التركيبية مقارنة بالنظرية الوصفية.

وإن هذا الوصف يقتضي تحديد مجموعة من السمات التعريفية لأنواع الجمل وأجزائها التركيبية، وكذلك السمات النحوية الوظيفية لوحدات جمل المدونة اللغوية عينة الدراسة في ضوء الغرض البحثي.

ينبغي أن نقف على المعلومات الأساسية التي سيقدمها التطبيق المنشود؛ للوقوف على المعطيات اللغوية المناسبة لمنهج البحث، وتعيين الأدوات التي تحقق الغاية منها. وقد شرع البحث في الفصل السابق في تعيين الأقسام الكلامية في ضوء التوصيف التركيبي الذي يستهدف التحليل النحوي الوظيفي.

ويتحدَّد هنا فئة العناصر الأساسية للتحليل التركيبي - التي ينتهجها البحث في عنونة العلاقات التركيبية - المتضمنة أنواع الجمل Clauses والمركبات Phrases، وهي العلامات الأكثر شهرة أو استقرارًا في التحليل التركيبي(المكوني)، وقد استخدمها بنك بنسلفانيا وبنك اللغة الصينية، وغيرهم مع مراعاة خصيصة كل لغة على حدة. يعرضها الجدول رقم ٣٣:

العناصر الأساسية للتحليل التركيبي	المصطلح العربي	المصطلح الإنجليزي	الاختصار
أنواع الجمل الصغرى	الجملة	Sentence	S
	الجملة شرطة	Sentence Bar	SBAR
	الجملة الاستفهامية	Sentence Question	SQ
	الجملة الاستفهامية شرطة	Sentence Question Bar	SBARQ
	الجملة الناقصة	Fragment	FRAG

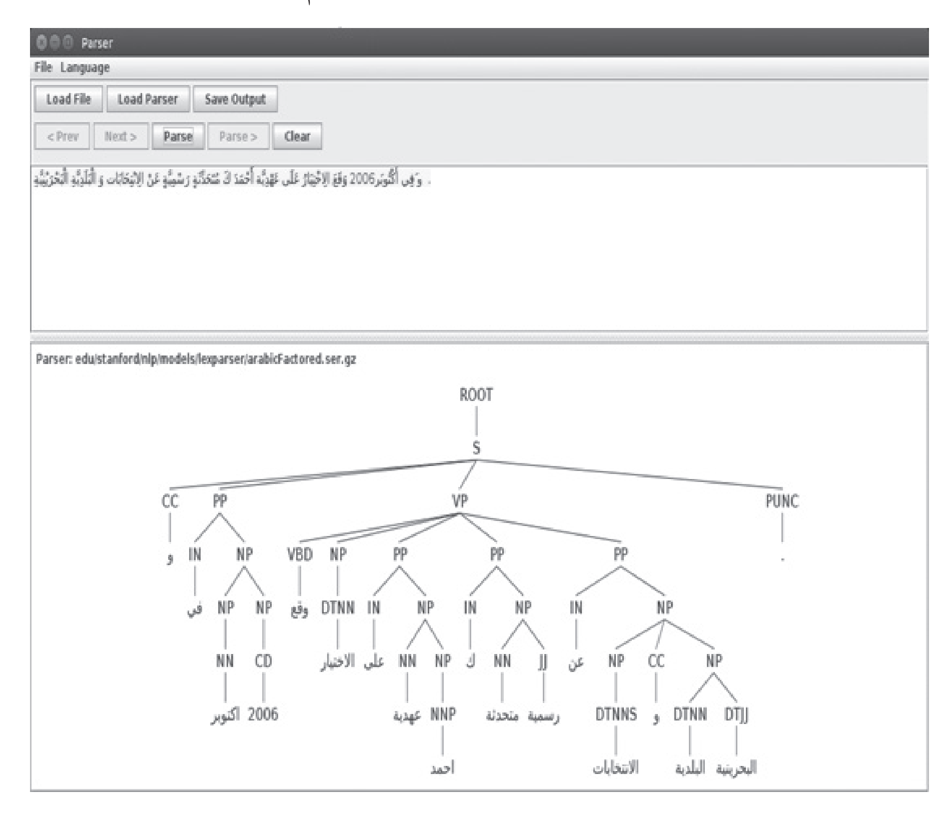
العناصر الأساسية للتحليل التركيبي	المصطلح العربي	المصطلح الإنجليزي	الاختصار
أنواع المركبات	المركب الاسمي	Noun Phrase	NP
	المركب الفعلي	Verb Phrase	VP
	المركب الحرفي	Preposition Phrase	PP
	المركب الظرفي	Adverb Phrase	ADVP
	المركب الوصفي	Adjective Phrase	ADJP
	المركب الموصول الاسمي	Wh-Noun Phrase	WHNP
	المركب الموصولي الظرفي	Wh Adverb Phrase -	WHADV
	المركب الموصولي الحرفي	Wh- Preposition Phrase	WHPP
	مركب العطف	Conj Phrase	CONJP
	التركيب الاعتراضي	Parenthetical Phrase	PRN
	مركب الأداة	Particle Phrase	PRT

الجدول: ٣٣ فئة السمات التركيبية

وقد استعانت الدراسة بآلية ستانفورد[1] للتحليل التركيبي Stanford Parser - كأداة مساعدة في عنونة نصوص المدونة اللغوية بالعلاقات التركيبية – التي تم تطويرها من قبل فريق معالجة اللغات الطبيعية بجامعة ستانفورد Stanford، والتي تتبع النهج الإحصائي في تحليل المدخلات، وقد حررت هذه الأداة بلغة الجافا البرمجية، مدعمة العديد من بيئات التشغيل المتعددة.

(١) هناك محللان تركيبان آخران، محلل بايكل Bikel التركيبي الذي استعملته مؤسسة شبكة البيانات اللغوية LDC في بناء مواردها النحوية ذي التمثيل المكوني، ومحلل بركلي Berkeley التابع لفريق معالجة اللغات الطبيعية بجامعة كاليفورنيا- بركلي، إلا أنهما يحتويان على النماذج النحوية للغة الإنجليزية Models of English syntax فحسب، ومن ثم يحتاجان إلى مدونة للتدريب Training Corpus في حالة بناء نموذج نحوي للغة أخرى.

ويعرض الشكل رقم ٤١ نموذجًا من تحليل آلية ستانفورد لجملة من نصوص المدونة اللغوية، وهي " و في أكتوبر عام ٢٠٠٦ وقع الاختيار على عهدية أحمد ك متحدثة رسمية عن الانتخابات والبلدية البحرينية". إسلام أون لاين ٢٠٠٨.



شكل: ٤١ آلية ستانفورد للتحليل التركيبي[١]

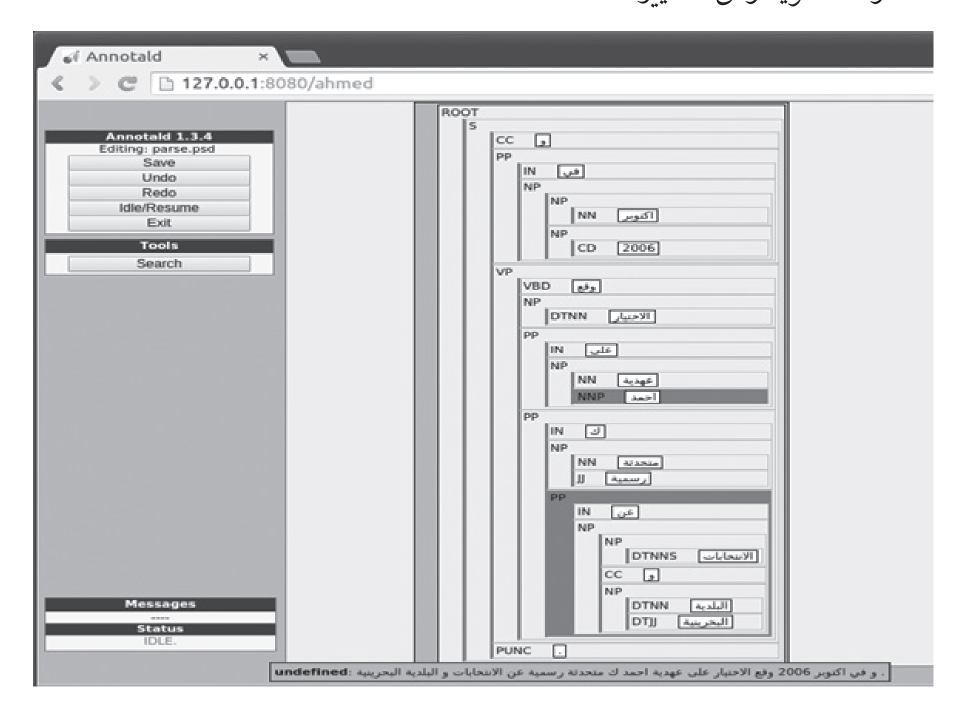
وقد استعانت الدراسة بأداة تحرير المدونات اللغوية الموصفة ذات الصيغة التقويسية Annotald التي تم تطويرها في الأصل من قبل اللغوي الأمريكي أنطون إنجسون Anton Ingason أثناء العمل على مشروع المدونة التاريخية الموصفة للغة الآيسلندية Icelandic Parsed Historical Corpus (IcePaHC) ، ثم تطورت لاحقا إلى ما صارت عليه الآن على يد اللغوية الأمريكية جانة بك Jana Beck. وقد خضعت هذه

(١) يمكن الاطلاع على آلية ستانفورد للتحليل التركيبي من خلال هذا الرابط:
Stanford-parser . (2014). http://nlp.stanford.edu/software/lex-parser.shtml

الأداة لرخصة جنو العامة GNU General Public License (GPL) التي يسمح بمقتضاها نسخ أو استخدام الأدوات باعتبارها ملكية عامة.[1]

وتم تحرير هذه الأداة كـ واجهة استخدام رسومية GUI باستخدام لغة الجافا سكريبت javascript، بحيث تمكِّن المستخدم من تحرير المدونة من خلال رؤية التحليلات نفسها، أما نواة النظام فقد حررت بلغة البايثون البرمجية.

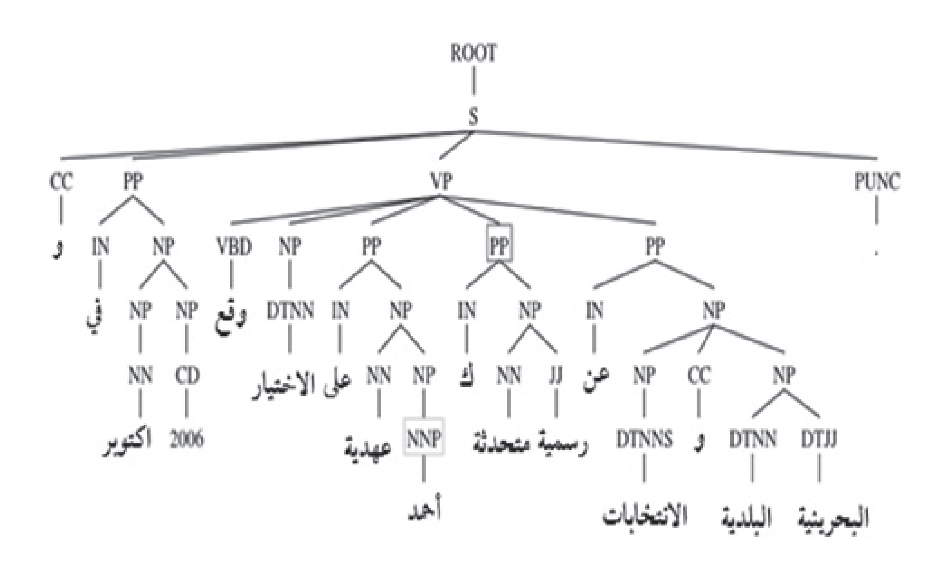
ويعرض الشكل رقم ٤٢ صورة للأداة أثناء تحرير مخرجات محلل ستانفورد التركيبي للمدونة اللغوية وفق المعايير المحددة.

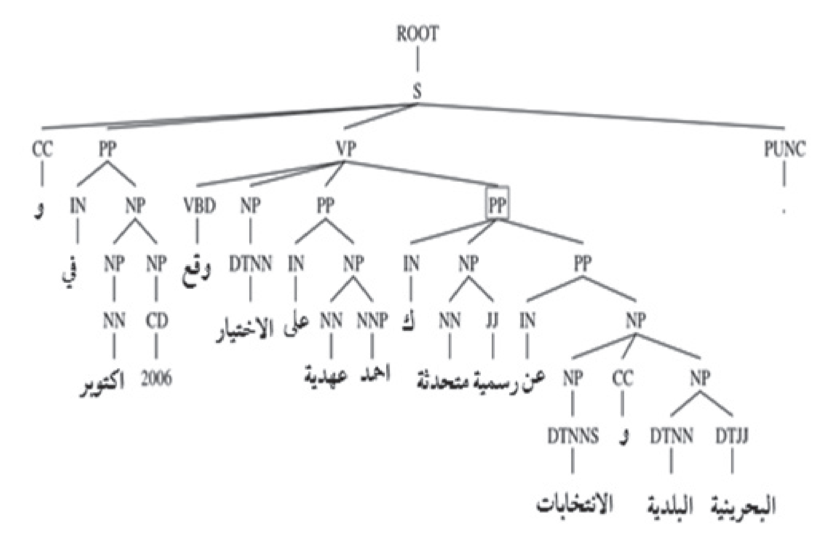


الشكل:٤٢ أداة Annotald لتحرير المدونات الموصفة تركيبيًا

بعد إجراء عملية التحرير على التقويس الشجري، يمكننا أن نختبر دقة الآلية من خلال المقارنة بين ما قبل التحرير وما بعد التحرير. ويوضح الشكلان ٤٣، ٤٤ نموذجًا لمواضع الاختلاف بين التحليلين.

(١) للاطلاع على الأداة، وتحميلها، من خلال هذا الرابط:
Annotald Program. (2015). https://annotald.github.io/





الشكل: ٤٣-٤٤ ما قبل التحرير وما بعد التحرير التركيبي

وبملاحظة التحليلين للجملة، نجد اختلافًا واضحًا بين التحليل الآلي والتحليل الذي أجرى عليه التعديلات، والتي تمثلت في توازي بنية المركب الاسمي NP ’’ عهدية أحمد‘‘ على اعتبار أنهما يشتركان في الوظيفة النحوية الدلالية، وانتقال المركب الحرفي PP من سيطرة المركب الفعلي مباشرة إلى سيطرة المركب الحرفي ’’ك متحدثة رسمية‘‘؛ لأنه ليس وسيطًا Argument للفعل.

ويمكن حساب دقة نتائج الآلية (التقويس) في المثال السابق مقارنة بالتعديلات التي أجريت عليه من خلال معادلتي الدقة والمراجعة:

الدقة Precision: هي حساب عدد الوحدات الصحيحة على قيمة عدد النتائج المسترجعة أي التي أخرجتها الآلية، وذلك باستخدام هذه المعادلة:

$$\text{Precision} = \frac{tp}{tp + fp}$$

حيث TP تعني True Positive القيم الإيجابية الصحيحة (عدد الأقواس الصحيحة)، وFP القيم الإيجابية الخاطئة (عدد الأقواس الخاطئة المثبتة في الخرْج) .

وبتطبيق المعادلة :

$$93.75 = \frac{15}{15 + 1} =$$

المراجعة Recall: هي حساب عدد الوحدات الصحيحة على قيمة عدد النتائج الكلية، باستخدام هذه المعادلة:

$$\text{Recall} = \frac{tp}{tp + fn}$$

حيث FN تعني False Negative القيم السلبية الخاطئة (عدد الأقواس الخاطئة المثبتة أو غير المثبتة) .

وبتطبيق المعادلة :

$$88.24 = \frac{15}{15 + 2} =$$

وباتحاد نتائج الدقة والمراجعة، نحصل على مقياس متوسط الحساب للنتائج من خلال معادلة الوسط التوافقي:

$$F = \frac{1}{\alpha \frac{1}{P} + (1-\alpha)\frac{1}{R}} = \frac{(\beta^2+1)PR}{\beta^2 P + R}$$

واختصارها إلى القياس الطبيعي:

$$F = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

وبتطبيق المعادلة:

$$\frac{93.75 * 88.24}{93.75 + 88.24} 2 =$$

$$90.91 = 2 * 93.75 * 88.24/(93.75 + 88.24) =$$

هذا بالنسبة لاختبار الآلية على مستوى الجملة الواحدة، أما اختبار الآلية الذي أُجري على ١٠٪ من حجم بنك بنسلفانيا العربي، فيعرضها الجدول رقم ٣٤:

النموذج التركيبي Syntax Model	طول الجملة Length	الدقة Precision	المراجعة Recall	المتوسط التوافقي F-measure
ستانفورد	70 كلمة	80.37	79.36	79.86
	عام	78.92	77.72	78.32

الجدول: ٣٤ قياس أداة محلل ستانفورد التركيبي[1]

(1) Green, S & D.Manning, C. (2010). Better Arabic Parsing: Baselines, Evaluations, and Analysis. In Proceedings of the 23rd International Conference on Computational Linguistics (Coling). Beijing.

ثم كانت الخطوة التالية باستخدام أداة التوازي Mapping بين الأقسام الكلامية التي ينتهجها البحث (فئة آر دي آي الكلامية)، والأقسام الكلامية التي تنتهجها آلية ستانفورد[1] للتحليل التركيبي؛ ليظهر المثال السابق كالآتي:

"وفي أكتوبر ٢٠٠٦ وقع الاختيار على عهدية أحمد كمتحدثة رسمية عن الانتخابات والبلدية البحرينية". إسلام أون لاين ٢٠٠٦.

```
(ROOT
  (S (NullPrefix+Conj+NullSuffix وَ)
    (PP (NullPrefix+Prepos+NullSuffix في)
      (NP
        (NP (NullPrefix+Noun+NoSARF+Masc+Single+NullSuffix أُكْتُوبَرَ))
        (NP (CD 2006))))
    (VP (NullPrefix+Active+Verb+Past+NullSuffix وَقَعَ)
      (NP (Definit+Noun+NounInfinit+NullSuffix اَلِاخْتِيَارُ))
      (PP (NullPrefix+Prepos+NullSuffix عَلَى)
        (NP (NullPrefix+Noun+RelAdj+Femin+Single عَهْدِيَّةِ)
          (NP (NullPrefix+Noun+NoSARF+NullSuffix أَحْمَدَ))))
      (PP (NullPrefix+Prepos+NullSuffix كَ)
        (NP (NullPrefix+Noun+SubjNoun+Femin+Single مُتَحَدِّثَةٍ) (NullPrefix+N
oun+RelAdj+Femin+Single رَسْمِيَّةٍ)))
      (PP (NullPrefix+Prepos+NullSuffix عَنْ)
        (NP
          (NP (Definit+Noun+NounInfinit+Plural+Femin اَلِانْتِخَابَاتِ))
          (NullPrefix+Conj+NullSuffix وَ)
          (NP (Definit+Noun+RelAdj+Femin+Single الْبَلَدِيَّةِ) (Definit+Noun+R
elAdj+Femin+Single الْبَحْرَيْنِيَّةِ)))))
    (PUNC .)))
```

(١) انظر للشكل رقم ٤٣-٤٤ للتعرف على الأقسام الكلامية التي تنتهجها آلية ستانفورد.

٤-١ فئة العناصر الأساسية للتحليل التركيبي

٤-١-١على مستوى التركيب الجملي

تعددت تصانيف أنواع الجمل لدى علماء النحو العربي القدامى والمعاصرين، إذ نجد ابن هشام يقسمها إلى ثلاثة أنواع، فيقول: «انقسام الجملة إلى اسمية وفعلية وظرفية، فالاسمية: هي التي صدرها اسم كزيد قائم، وهيهات العقيق، وقائم الزيدان عند من جوزه وهو الأخفش والكوفيون، والفعلية: هي التي صدرها فعل كـ «قام زيد، وضُرب اللص، وكان زيد قائما، ويقول زيد، وقمْ»، والظرفية المصدرة بظرف أو مجرور نحو: أعندك زيد؟ أو: أفي الدار زيد؟ إذا قدرت زيدًا فاعلا بالظرف والجار والمجرور.[١]

أمَّا أبو عليّ الفارسي، والزمخشري وعبد القاهر الجرجاني فيصنفوها إلى أربعة أنواع، إذ يقول عبد القاهر:«فقد حصل لك أربعة أضرب من الجمل، وهي في الأصل اثنتان: الجملة من الفعل والفاعل، والجملة من المبتدأ والخبر»[٢]، بينما الشائع عند النحويين أن الجملة نوعان: اسمية متمثلة في عنصري الإسناد المبتدأ والخبر، وجملة فعلية متمثلة في عنصري الإسناد الفعل والفاعل.

وتوسع بعض النحويين في تصنيف ذلك التقسيم بجعله في إطار أكثر شمولية، حيث قسموا الجملة إلى جملة كبرى وصغرى، وقال:«الكبرى: هي الاسمية التي خبرها جملة نحو: «زيد قائم أبوه»، و«زيد أبوه قائم»، والصغرى: هي المبنية على المبتدأ كالجملة المخبر بها في المثالين»[٣]. وهو ما يتوافق مع تصنيف الدراسات اللغوية المعاصرة مع اختلاف التقسيمات الداخلية.[٤]

والواقع أن توخي أيًّا من هذه التصنيفات مشروط بمنهجية محددة، فليست كل هذه التصنيفات ملائمة لكل اللغات، إذ لكل لغة خصائصها التركيبية التي تمتاز بها عن غيرها، فخواص لغة الأدب التركيبية تختلف عن خواص لغة الصحافة التي اكتسبتها من مصادر عدة.

---

(١) الأنصاري (ابن هشام): مغني اللبيب عن كتب الأعاريب، تحقيق محمد محيي الدين عبد الحميد، المكتبة العصرية - بيروت، ١٩٩١، الجزء الثاني، ص ٤٣٣.

(٢) الجرجاني (عبد القاهر): كتاب المقتصد في شرح الإيضاح، تحقيق د. كاظم بحر المرجان، دار الرشيد للنشر - الجمهورية العراقية ١٩٨٢، مجلد ١، ص ٢٧٧.

(٣) الأنصاري (ابن هشام): السابق، ص ٤٣٧.

(٤) للاطلاع حول بعض تصنيفات الجمل في الدراسات اللغوية المعاصرة، ينظر: البهنساوي (حسام): القواعد التحويلية في ديوان حاتم الطائي، ص ٨٧-٨٨.

ولما كانت الدراسة قد اتخذت الإسناد والتركيب التام المفيد وسيلة لتحديد أبعاد الجملة في تجزئتها لنصوص المدونة، فقد انحصرت أنواع الجملة في أربعة أنواع:

- **الجملة البسيطة**: هي الجملة المكونة من عملية إسنادية مستغنية بنفسها عن غيرها شكلًا ودلالة، وهي اسمية وفعلية[١].
- **الجملة المركبة**: هي الجملة المكونة من مركبين أحدهما مرتبط بالآخر، ومتوقف عليه، يتم به الكلام.
- **الجملة الناقصة**: هي الجملة التي تفقد أحد أركانها الإسنادية، وتتمثل في عناوين الصحف، والقصص،....
- **الجملة التكميلية**: هي الجملة المكونة من عناصر داخلة أو ملازمة لعناصر الإسناد الأساسية، وتتصدر بمكمل complementizer مثل ضمائر الوصل.[٢]

وتتضح هذه الأنواع في ضوء تمثيلها في تلك العلامات:

- الجملة "S"

تستخدم هذه العلامة لنمطي الجملة البسيطة والمركبة بنوعيها الاسمي والفعلي، ممثلة لرأس الجملة.

"لم يسقط السادات هذه المرة في فخ النمط."

```
(ROOT
  (S
    (VP
      (PRT (NullPrefix+JAAZIMA+NullSuffix لم))
      (Present+Active+Verb+NullSuffix يسقط)
      (NP (Definit+Noun+Femin+Plural السادات))
      (NP
        (NP (NullPrefix+Noun+DemoPro+Femin+Single هذه))
        (NP (Definit+Noun+Femin+Single المرة)))
      (PP (NullPrefix+Prepos+NullSuffix في)
        (NP (NullPrefix+Noun+NullSuffix فخ)
          (NP (Definit+Noun+NullSuffix النمط)))))
    (PUNC_ .)))
```

(١) عاشور (المنصف): بنية الجملة العربية بين التحليل والنظرية، ص ٥٣.

(2) Radford, A. (2009). An Introduction to English Sentence Structure. Cambridge University Press. P.49.

«معروضات المتحف توحي بأن الإنسان عاش في زمن الديناصورات.»

```
(ROOT
  (S
    (NP (NullPrefix+Noun+ObjNoun+Femin+Plural معروضات)
      (NP (Definit+Noun+NullSuffix المتحف)))
    (VP (Present+Active+Verb+NullSuffix توحي)
      (PP (NullPrefix+Prepos+NullSuffix ب)
        (SBAR (NullPrefix+ParticleNAASIKH+NullSuffix أن)
          (S
            (NP (Definit+Noun+NullSuffix الإنسان))
            (VP (NullPrefix+Active+Verb+Past+NullSuffix عاش)
              (PP (NullPrefix+Prepos+NullSuffix في)
                (NP (NullPrefix+Noun+NullSuffix زمن)
                  (NP (Translit الديناصورات))))))))
    (PUNC_ .)))
```

"نحن لدينا كل الثقة بأنفسنا."

```
(ROOT
  (S
    (NP (NullPrefix+Noun+SubjPro+NullSuffix نحن))
    (NP (NullPrefix+DZARF+Adjunct+NullSuffix لدي)
      (NP (NullPrefix+Noun+Pronoun+NullSuffix نا)))
    (NP
      (NP (NullPrefix+Noun+NullSuffix كل)
        (NP (Definit+Noun+NounInfinit+Femin+Single الثقة)))
      (PP (NullPrefix+Prepos+NullSuffix_CON ب)
        (NP (NullPrefix+Noun+Plural+NullSuffix أنفس)
          (NP (NullPrefix+Noun+Pronoun+NullSuffix نا)))))
    (PUNC_ .)))
```

"وقد أكدت الهيئات المنظمة للمؤتمر على استمرار دعمها للكيانات الجديدة الناشئة؛ حيث ستقوم بتوفير ما ستحتاجه تلك الكيانات من تدريب لأعضائها، وعقد فاعليات تساعد على تناقل الخبرات بين أعضائها."

```
(ROOT
  (S
    (S (NullPrefix+Conj+NullSuffix و)
      (VP
        (PRT (NullPrefix+NoSyntaEffect+NullSuffix قد))
        (NullPrefix+Verb+Past+Femin+Single أكدت)
        (NP
          (NP (Definit+Noun+Femin+Plural الهيئات) (Definit+Noun+SubjNoun
+Femin+Single المنظمة))
          (PP (NullPrefix+Prepos+NullSuffix ل)
            (NP (Definit+Noun+ObjNoun+NullSuffix المؤتمر))))
        (PP (NullPrefix+Prepos+NullSuffix على)
          (NP (NullPrefix+Noun+NounInfinit+NullSuffix استمرار)
            (NP
```

```
               (NP (NullPrefix+Noun+NullSuffix دعم)
                 (NP (NullPrefix+Noun+Pronoun+NullSuffix ها)))
               (PP (NullPrefix+Prepos+NullSuffix ل)
                 (NP
                   (NP (Definit+Noun+Femin+Plural الكيانات) (Definit+No
un+ExaggAdj+Femin+Single الجديدة) (Definit+Noun+SubjNoun+Femin+Single
الناشئة)))))))))
     (PUNC_ ;)
     (SBAR
       (WHADVP (NullPrefix+CondJAAZIMA+DZARF+NullSuffix حيث))
       (S
         (VP (Future+Present+Active+Verb+NullSuffix ستقوم)
           (PP (NullPrefix+Prepos+NullSuffix ب)
             (NP (NullPrefix+Noun+NounInfinit+NullSuffix توفير)
               (SBAR (NullPrefix+Noun+RelPro+NullSuffix ما)
                 (S
                   (VP (Future+Present+Active+Verb+NullSuffix ستحتاج)
                     (NP (NullPrefix+Noun+Pronoun+NullSuffix ه))
                     (NP
                       (NP (NullPrefix+Noun+DemoPro+Femin+Single تلك))
                       (NP (Definit+Noun+Femin+Plural الكيانات)))
                     (PP (NullPrefix+Prepos+NullSuffix من)
                       (NP
                         (NP (NullPrefix+Noun+NounInfinit+NullSuffix
تدريب))
                         (PP (NullPrefix+Prepos+NullSuffix ل)
                           (NP (NullPrefix+Noun+Plural+NullSuffix أعضاء)
                             (NP (NullPrefix+Noun+Pronoun+NullSuffix
ها))))))
                     (PUNC_ ,) (NullPrefix+Conj+NullSuffix و)
                     (S
                       (NP (NullPrefix+Noun+NullSuffix عقد)
                         (NP (NullPrefix+Noun+SubjNoun+Plural+RelAdj+Fem
in فاعليات)))
                       (VP (Present+Active+Verb+NullSuffix تساعد)
                         (PP (NullPrefix+Prepos+NullSuffix على)
                           (NP (NullPrefix+Noun+NounInfinit+NullSuffix
تناقل)
                             (NP (Definit+Noun+Femin+Plural الخبرات))))
                         (NP (NullPrefix+Noun+NullSuffix بين)
                           (NP (NullPrefix+Noun+Plural+NullSuffix أعضاء)
                             (NP (NullPrefix+Noun+Pronoun+NullSuffix
ها))))))))))))))
     (PUNC_ .)))
```

"الإسلاميون يتراجعون بعد تشقق حركة الإصلاح قبل أن تعلن وزارة الداخلية الجزائرية عن القوائم الانتخابية المقبولة وفق قانون الانتخابات."

```
(ROOT
  (S
     (NP (Definit+Noun+NounInfinit+Plural+MARF+RelAdj+Masc+NonAdjunct
الإسلاميون))
     (VP (Present+Active+Verb+MARF+SubjPro يتراجعون)
       (NP (NullPrefix+Noun+NullSuffix بعد)
         (NP (NullPrefix+Noun+NounInfinit+NullSuffix تشقق)
           (NP (NullPrefix+Noun+Femin+Single حركة)
             (NP (Definit+Noun+NounInfinit+NullSuffix الإصلاح)))))
       (NP (NullPrefix+Noun+NullSuffix قبل)
```

```
        (SBAR (NullPrefix+ParticleNAASSIB+NullSuffix أن)
          (S
            (VP (Present+Active+Verb+NullSuffix تعلن)
              (NP (NullPrefix+Noun+Femin+Single وزارة)
                (NP (Definit+Noun+SubjNoun+RelAdj+Femin+Single
الداخلية) (Definit+Noun+NoSARF+Plural+RelAdj+Femin+Single الجزائرية)))
              (PP (NullPrefix+Prepos+NullSuffix عن)
                (NP (Definit+Noun+NoSARF+Plural+NullSuffix القوائم) (Defi
nit+Noun+NounInfinit+RelAdj+Femin+Single الانتخابية) (Definit+Noun+ObjNou
n+Femin+Single المقبولة)))
              (NP (NullPrefix+Noun+NullSuffix وفق)
                (NP (NullPrefix+Noun+NullSuffix قانون)
                  (NP (Definit+Noun+NounInfinit+Femin+Plural
الانتخابات))))))))
    (PUNC_ .)))
```

● الجملة "شرطة" "SBAR"

تستخدم هذه العلامة للجملة التكميلية المتمثلة في جملة الصلة التي تتصدرها اسم موصول (الرأس) سواء أكان خاصًا أو مشتركًا، والجملة الثانوية مثل جملة مقول القول.[1]

"فهي الغامضة، التي لا تريد استعراض قلبها وبضاعتها النفسية، من أول لحظة."

```
(ROOT
  (S
    (PRT (NullPrefix+Conj+NullSuffix ف))
    (NP
      (NP
        (NP (NullPrefix+Noun+SubjPro+NullSuffix هي))
        (ADJP (Definit+Noun+SubjNoun+Femin+Single الغامضة)))
      (PUNC_ ,)
      (SBAR
        (WHNP (Definit+Noun+RelPro+Femin+Single التي))
        (S
          (VP
            (PRT (NullPrefix+LAA+NullSuffix لا))
            (Present+Active+Verb+NullSuffix تريد)
            (NP
              (NP (NullPrefix+Noun+NounInfinit+NullSuffix استعراض)
                (NP (NullPrefix+Noun+NullSuffix قلب)
                  (NP (NullPrefix+Noun+Pronoun+NullSuffix ها))))
              (NullPrefix+Conj+NullSuffix و)
              (NP
                (NP
                  (NP (NullPrefix+Noun+Femin+Single بضاعة))
                  (NP (NullPrefix+Noun+ProNoun+NullSuffix ها)))
                (ADJP (Definit+Noun+RelAdj+Femin+Single النفسية))))))))
    (PUNC_ ,)
    (PP (NullPrefix+Prepos+NullSuffix من)
      (NP
        (ADJP (NullPrefix+Noun+Masc+NullSuffix أول)
          (NP (NullPrefix+Noun+Femin+Single لحظة)))))
    (PUNC_ .)))
```

(1) Maamouri, M. et al,. (2011). Penn Treebank Guidlines. Linguistic Data Consorttium. P.21.

"وقال في حواره مع إسلام أون لاين: إن السينما السعودية ما زالت في البداية."

```
(ROOT
  (S (NullPrefix+Conj+NullSuffix و)
    (VP (NullPrefix+Active+Verb+Past+NullSuffix قال)
      (PP (NullPrefix+Prepos+NullSuffix في)
        (NP
          (NP (NullPrefix+Noun+NullSuffix حوار)
            (NP (NullPrefix+Noun+Pronoun+NullSuffix هـ)))
          (NP (NullPrefix+DZARF+Adjunct+NullSuffix مع)
            (NP (Translit إسلام-أونلاين)))))
      (PUNC_ :)
      (SBAR (NullPrefix+ParticleNAASIKH+NullSuffix إن)
        (S
          (NP (Translit السينما) (Definit+Noun+RelAdj+Femin+Single
السعودية))
          (VP
            (PRT (NullPrefix+NoSyntaEffect+NullSuffix ما))
            (NullPrefix+Active+Verb+Past+Femin+Single زالت)
            (PP (NullPrefix+Prepos+NullSuffix في)
              (NP (Definit+Noun+Femin+Single البداية)))))))
    (PUNC_ .)))
```

• الجملة الاستفهامية "SQ"

تستخدم هذه العلامة للجملة الاستفهامية التي تشتمل على حرف استفهام (همزة – هل)، والتي يتوقع منها الإجابة المباشرة بلا أو نعم أو بلى.[1]

"هل تتخوفون من وقوع حرب أهلية؟"

```
(ROOT
  (SQ
    (PRT (NullPrefix+InterrogArticle+NullSuffix هل))
    (VP (Present+Active+Verb+MARF+SubjPro تتخوفون)
      (PP (NullPrefix+Prepos+NullSuffix من)
        (NP (NullPrefix+Noun+NullSuffix وقوع)
          (NP (NullPrefix+Noun+NullSuffix حرب)
(NullPrefix+Noun+RelAdj+Femin+Single أهلية)))))
    (PUNC_ ?)))
```

(1) Maamouri, M. et al,. (2011). Ibid. P.23.

• الجملة الاستفهامية "شرطة" "SQBAR"

تستخدم هذه العلامة للجملة الاستفهامية التي تشتمل على اسم استفهام، وغالبا ما يتوقع منها إجابة ضمنية.[1]

"كيف تنظر إلى المؤسسة الدينية الرافضة للسينما؟"

```
(ROOT
  (SBARQ
    (WHADVP (NullPrefix+Noun+NullSuffix كيف))
    (S
      (VP (Present+Active+Verb+NullSuffix تنظر)
        (PP (NullPrefix+Prepos+NullSuffix إلى)
          (NP
            (NP (Definit+Noun+ObjNoun+Femin+Single المؤسسة) (Definit+N
oun+RelAdj+Femin+Single الدينية) (Definit+Noun+SubjNoun+Femin+Single
الرافضة))
            (PP (NullPrefix+Prepos+NullSuffix ل)
              (NP (Translit السينما)))))))
    (PUNC_ ?)))
```

• الجملة الناقصة "FRAG"

تستخدم هذه العلامة للجملة التي تفقد أحد أركانها الإسنادية، وغالبا ما تتمثل في عناوين الصحف، والقصص، والقوائم.[2]

"مغازلة انتخابية."

```
(ROOT
  (FRAG
    (NP (NullPrefix+Noun+Femin+Single مغازلة)
      (NP (NullPrefix+Noun+NounInfinit+RelAdj+Femin+Single انتخابية)))
    (PUNC_ .)))
```

"اصطدام مع داروِن."

```
 (ROOT
  (FRAG
    (NP (NullPrefix+Noun+NounInfinit+NullSuffix اصطدام))
    (NP (NullPrefix+DZARF+Adjunct+NullSuffix مع)
      (NP (Translit داروِن)))
    (PUNC_ .)))
```

"خامسا: ظاهرة الشركات العائلية."

```
(ROOT
  (FRAG
    (NP (NullPrefix+Noun+SubjNoun+NullSuffix خامسا))
    (PUNC_ :)
    (NP (NullPrefix+Noun+SubjNoun+Femin+Single ظاهرة)
      (NP (Definit+Noun+Femin+Plural الشركات) (Definit+Noun+SubjNoun+Rel
Adj+Femin+Single العائلية)))
    (PUNC_ .)))
```

(1) Maamouri, M. et al,. (2011). Ibid. P.25.

(2) Maamouri, M. et al,. (2011). Ibid. P.27.

٤-١-٢ على مستوى المركبات

● المركب الاسمي "NP"

يقصد به الهيئة التركيبية المبدوءة باسم ليس مشتقا عاملا عمل فعله، وليس مصدرًا عاملا عمل فعله، ويمثل رأس المركب في التركيب الشجري.[(١)]

"ونحن الآن لا نتوفر على إحصائيات دقيقة."

```
(ROOT
  (S (NullPrefix+Conj+NullSuffix و)
    (NP (NullPrefix+Noun+SubjPro+NullSuffix نحن))
    (NP (Definit+Noun+NullSuffix الآن))
    (VP
      (PRT (NullPrefix+LAA+NullSuffix لا))
      (Present+Active+Verb+NullSuffix نتوفر)
      (PP (NullPrefix+Prepos+NullSuffix على)
        (NP (NullPrefix+Noun+NounInfinit+Plural+RelAdj+Femin إحصائيات)
(NullPrefix+Noun+ExaggAdj+Femin+Single دقيقة))))
    (PUNC_ .)))
```

● المركب الفعلي "VP"

يقصد به الهيئة التركيبية المبدوءة بفعل أو مصدر عامل عمل فعله، ويمثل رأس المركب في التركيب الشجري.[(٢)]

"وتستمر حتى ٢٤-٢-٢٠٠٨."

```
(ROOT
  (S (NullPrefix+Conj+NullSuffix و)
    (VP (Present+Active+Verb+NullSuffix تستمر)
      (PP (NullPrefix+Prepos+ParticleNAASSIB+NullSuffix حتى)
        (NP (CD 24-2-2008))))
    (PUNC_ .)))
```

● المركب الحرفي "PP"

يقصد به الهيئة التركيبية المبدوءة بحرف من حروف الجر سواء أكان أصليًّا أو زائدًا، ويمثل رأس المركب في التركيب الشجري.[(٣)]

"عجزوا عن صنع جيل من الرائدات النسويات."

---

(١) ينظر: عبادة (محمد إبراهيم): الجملة العربية، مكوناتها-أنواعها-تحليلها، ص ٥٧.

(٢) ينظر: عبادة (محمد إبراهيم): المصدر السابق، ص ٤٤.

(٣) ينظر: عبادة (محمد إبراهيم): المصدر السابق، ص ١٢٤.

```
(ROOT
  (S
    (VP (NullPrefix+Active+Verb+Past+SubjPro عجزوا)
      (PP (NullPrefix+Prepos+NullSuffix عن)
        (NP (NullPrefix+NounNounInfinit+NullSuffix صنع)
          (NP
            (NP (NullPrefix+Noun+NullSuffix جيل))
            (PP (NullPrefix+Prepos+NullSuffix من)
              (NP (Definit+Noun+SubjNoun+Femin+Plural الرائدات) (Definit
+Noun+Plural+RelAdj+Femin النسويات)))))))
    (PUNC_ .)))
```

• المركب الظرفي "ADVP"

يقصد به الهيئة التركيبية المبدوءة بالظرف الدال على زمان ومكان إنجاز الحدث، والمتضمن معنى " في"، والذي يمثل رأس المركب في التركيب الشجري.[(١)]

"فهناك ٦٠٪ من أفراد العينة حاصلون على شهادة جامعية."

```
(ROOT
  (S (NullPrefix+Conj+NullSuffix ف)
    (ADVP (NullPrefix+DZARF+NullSuffix هناك))
    (NP
      (NP (CD 60%))
      (PP (NullPrefix+Prepos+NullSuffix من)
        (NP (NullPrefix+Noun+Plural+NullSuffix أفراد)
          (NP (Definit+Noun+Femin+Single العينة)))))
    (NP (NullPrefix+Noun+SubjNoun+Plural+MARF+Masc+NonAdjunct حاصلون)
      (PP (NullPrefix+Prepos+NullSuffix على)
        (NP (NullPrefix+Noun+NounInfinit+Femin+Single شهادة) (NullPrefix+
Noun+SubjNoun+RelAdj+Femin+Single جامعية))))
    (PUNC_ .)))
```

• المركب الوصفي "ADJP"

يقصد به الهيئة التركيبية المبدوءة بوصف أو مشتق محض (اسم فاعل، اسم مفعول، صفة مشبهة، اسم تفضيل)، ممثلا رأس المركب في التركيب الشجري.[(٢)]

"وبقول آخر: الدور هو حصيلة جمع الممثل بالشخصية."

(١) ينظر: عبادة (محمد إبراهيم): المصدر السابق، ص ١١٦.

(٢) ينظر: عبادة (محمد إبراهيم): المصدر السابق، ص ٨٢.

```
(ROOT
  (S
    (S (NullPrefix+Conj+NullSuffix و)
      (PP (NullPrefix+Prepos+NullSuffix ب)
        (NP
          (NP (NullPrefix+Noun+NullSuffix قول))
          (ADJP (NullPrefix+Noun+NoSARF+NullSuffix آخر)))))
    (PUNC_ :)
    (S
      (NP (Definit+Noun+NullSuffix الدور))
      (NP (NullPrefix+Noun+SubjPro+NullSuffix هو))
      (NP (NullPrefix+Noun+ExaggAdj+Femin+Single حصيلة)
        (NP (NullPrefix+Noun+NullSuffix جمع)
          (NP (Definit+Noun+SubjNoun+NullSuffix الممثل))))
      (PP (NullPrefix+Prepos+NullSuffix ب)
        (NP (Definit+Noun+RelAdj+Femin+Single الشخصية))))
    (PUNC_ .)))
```

• المركب الموصول الاسمي "WHNP"

يقصد به الهيئة التركيبية المبدوءة باسم موصول، والاسم الموصول هو ما لايصير جزءا من جملة إلا بصلة وعائد، ويمثل رأس المركب في التركيب الشجري[1]

"وهو ما يجعل طبيعة الاستثمار أقرب إلى المقامرة."

```
(ROOT
  (S (NullPrefix+Conj+NullSuffix و)
    (NP (NullPrefix+Noun+SubjPro+NullSuffix هو))
    (SBAR
      (WHNP (NullPrefix+Noun+RelPro+NullSuffix ما))
      (S
        (VP (Present+Active+Verb+NullSuffix يجعل)
          (NP (NullPrefix+Noun+ExaggAdj+Femin+Single طبيعة)
            (NP (Definit+Noun+NounInfinit+NullSuffix الاستثمار)))
          (ADJP (NullPrefix+Noun+NoSARF+NullSuffix أقرب)
            (PP (NullPrefix+Prepos+NullSuffix إلى)
              (NP (Definit+Noun+Femin+Single المقامرة)))))))
    (PUNC_ .)))
```

• المركب الموصول الظرفي "WHADV"

يقصد به الهيئة التركيبية المبدوءة بمصدرية ظرفية (كلما، بينما، حينما،...) ، تمثل رأس المركب في التركيب الشجري.

"تعلم أنه كلما ارتفعت سوق المال، اقتربت لحظة الانخفاض، والعكس صحيح."

(١)ينظر: عبادة (محمد إبراهيم): المصدر السابق، ص ١٠١.

```
(ROOT
  (S
    (S
      (VP (Present+Imperative+Active+Verb+NullSuffix تعلم)
        (SBAR (NullPrefix+ParticleNAASIKH+NullSuffix أن)
          (S
            (NP (NullPrefix+Noun+Pronoun+NullSuffix ـه))
            (SBAR
              (WHADVP (NullPrefix+CondNotJAAZIMA+NullSuffix كلما))
              (S
                (VP (NullPrefix+Verb+NullSuffix ارتفع)
                  (NP (NullPrefix+Noun+NullSuffix سوق)
                    (NP (Definit+Noun+NullSuffix المال))))))
            (PUNC_ ,)
            (VP (NullPrefix+Verb+Past+Femin+Single اقتربت)
              (NP (NullPrefix+Noun+Femin+Single لحظة)
                (NP (Definit+Noun+NounInfinit+Intransitive+NullSuffix
الانخفاض))))))))
    (PUNC_ ,) (NullPrefix+Conj+NullSuffix و)
    (S
      (NP (Definit+Noun+NullSuffix العكس))
      (ADJP (NullPrefix+Noun+ExaggAdj+NullSuffix صحيح)))

    (PUNC_ .)))
```

• مركب عطف "CONJP"

يقصد به المكون التركيبي المبدوء بحرف عطف (عادة ما يكون حرف الواو)، عطفاً على مركب آخر.[1]

"ولكن هل يوجد في نطاق التعامل مع الأزمة القائمة ما يوحي بسواه؟"

```
(ROOT
  (SQ
    (CONJP (NullPrefix+Conj+NullSuffix و) (NullPrefix+NoSyntaEffect+NullS
uffix لكن))
    (PRT (NullPrefix+InterrogArticle+NullSuffix هل))
    (VP (Present+Passive+Verb+NullSuffix يوجد)
      (PP (NullPrefix+Prepos+NullSuffix في)
        (NP (NullPrefix+Noun+NullSuffix نطاق)
          (NP
            (NP (Definit+Noun+NounInfinit+NullSuffix التعامل))
            (NP (NullPrefix+DZARF+Adjunct+NullSuffix مع)
              (NP (Definit+Noun+Femin+Single الأزمة) (Definit+Noun+SubjNo
un+Femin+Single القائمة))))))
      (SBAR
        (WHNP (NullPrefix+Noun+RelPro+NullSuffix ما))
        (S
          (VP (Present+Active+Verb+NullSuffix يوحي)
            (PP (NullPrefix+Prepos+NullSuffix ب)
              (NP (NullPrefix+Noun+NullSuffix سوى)
                (NP (NullPrefix+Noun+Pronoun+NullSuffix ـه)))))))))
    (PUNC_ ?)))
```

(1) Maamouri, M. et al,. (2011). Ibid. P.59.

هناك علامات أخرى تدخل في إطار تحديد ماهية المركب:

- التركيب الاعتراضي PRN

يراد بهذه العلامة الدلالة على التركيب الاعتراضي الموصول بالإسناد والتلفظ، فلا يحسن السكوت إلا به، ويستعمل الاعتراض للتفخيم والتأكيد وتحسين الكلام وتسديده.[(١)]

"فالمسلم الحق (ذكرا وأنثى) هو الذي يسلم أموره إلى عبد الرحمن."

```
(ROOT
  (S
    (PRT (NullPrefix+Conj+NullSuffix ف))
    (NP
      (NP
        (NP (Definit+Noun+SubjNoun+NullSuffix المسلم))
        (NP (Definit+NounNounInfinit+NullSuffix الحق)))
      (PUNC_ -)
      (PRN
        (NP (NullPrefix+Noun+NullSuffix ذكرا))
        (NullPrefix+Conj+NullSuffix و)
        (NP (NullPrefix+Noun+NullSuffix أنثى)))
      (PUNC_ -))
    (S
      (NP (NullPrefix+Noun+SubjPro+NullSuffix هو))
      (SBAR
        (WHNP (Definit+Noun+RelPro+Masc+Single+NullSuffix الذي))
        (S
          (VP (Present+Active+Verb+NullSuffix يسلم)
            (NP (NullPrefix+Noun+NullSuffix أمور)
              (NP (NullPrefix+Noun+Pronoun+NullSuffix ـه)))
            (PP (NullPrefix+Prepos+NullSuffix إلى)
              (NP (NullPrefix+Noun+NullSuffix عبد-الرحمن)))))))

    (PUNC_ .)))
```

(١) ابن هشام (الأنصاري): السابق، ج٢ ص ٣٩٢. ينظر أيضا : عاشور (المنصف): بنية الجملة العربية بين التحليل والنظرية، ص ٢٣١.

- مركب الأداة PRT

يراد بهذه العلامة الدلالة على الأدوات التي تدخل على رأس المركبات لاسيما المركب الفعلي والاسمي.

"لا يتجاوز زمنه."

```
(ROOT
  (S
    (VP
      (PRT (NullPrefix+LAA+NullSuffix لا))
      (Present+Active+Verb+NullSuffix يـتـجـاوز)
      (NP (NullPrefix+Noun+NullSuffix زمـن)
        (NP (NullPrefix+Noun+Pronoun+NullSuffix ـه))))
    (PUNC_ .)))
"وهـذا ذنـبـه بـلا شك."

(ROOT
  (S (NullPrefix+Conj+NullSuffix و)
    (NP (NullPrefix+Noun+DemoPro+Masc+Single+NullSuffix هذا))
    (NP
      (NP (NullPrefix+Noun+NullSuffix ذنـب)
        (NP (NullPrefix+Noun+Pronoun+NullSuffix ـه)))
      (PP (NullPrefix+Prepos+NullSuffix ب)
        (NP
          (PRT (NullPrefix+LAA+NullSuffix لا))
          (NullPrefix+Noun+NullSuffix شك))))

    (PUNC_ .)))
```

٤-٢ فئة السمات النحوية الوظيفية

تحاول الدراسة أن تشمل هذه الفئة جميع السمات النحوية الوظيفية التي تساعد على إدراك العلاقات الأساسية بين وحدات الجملة، ومن ثم فهم معناها، وذلك باعتبارها الخطوة الأساسية في التحليل الدلالي، الذي يتم بموجبه الفهم الآلي للغات الطبيعية المتمثل في عدد من التطبيقات اللغوية الحاسوبية كاسترجاع المعلومات Information Retrieval وإجابة الأسئلة Question answering وغيرها.

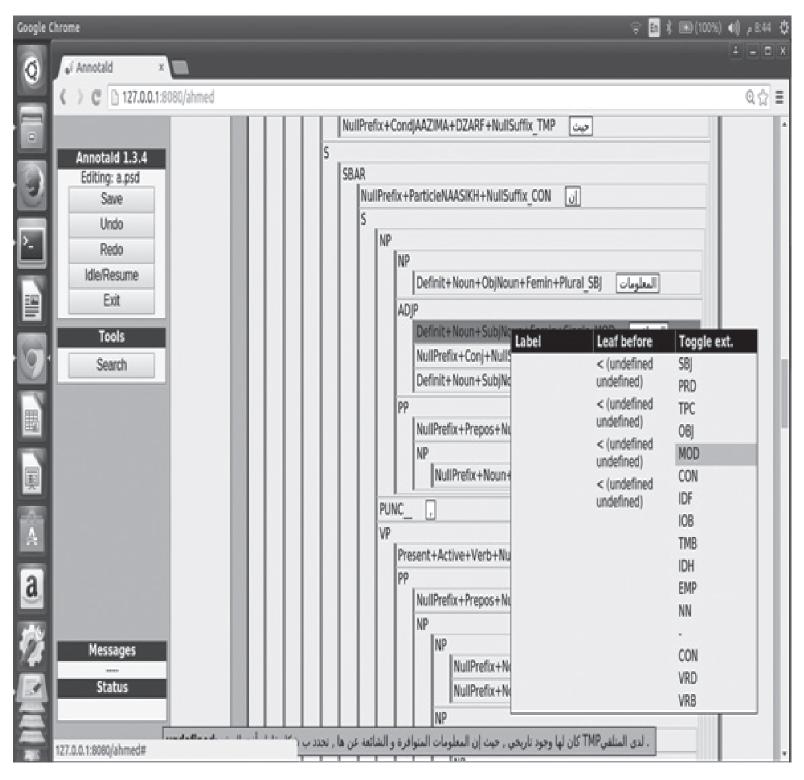
وقد حاول الباحث أن يستخلص جميع صور العلاقات النحوية الدلالية - من مصادر الكتب النحوية واللغوية - في أقل عدد من الرموز، بحيث تتناسب فيما بعد مع ما تقضتيه سبل تقنيات الذكاء الاصطناعي.

ويعرض الجدول رقم ٣٥ فئة هذه السمات جملة، ثم يتبعها تفصيلا موضحًا بالأمثلة المقتبسة من المدونة اللغوية التي طبقت عليها.

الفئة	المصطلح العربي	المصطلح الإنجليزي	الاختصار
الوظائف النحوية Grammatical functions	-المسند	Subject	SBJ
	-المسند إليه	Predicate	PRD
	-الموضوع	Topic	TPC
الوظائف النحوية Grammatical functions	-المفعولية	Object	OBJ
	-فعل	Verb	VRB
	-تمييز	Tamyiyz	TMZ
	-وصف	Modifier	MOD
	-إضافة	IDafa	IDF
	-جار ومجرور	Indirect object	IOB
	-التوكيد	emphasis	EMP
	-الزمان والمكان	Locative Temporal	TMB
	-السببية	Purposive	PRB
	-الإيضاح	IDah	IDH
	-المتلازم	compund	NN
	-فعل_فاعل	Verb_SBJ	VRD
	-ربط	Conjunction	CON
	-علامة	For punc and sym	MRK

الجدول:٣٥ فئة السمات النحوية الوظيفية

وقد استعانت الدراسة بأداة تحرير المدونات اللغوية الموصفة Annotald في عنونة المدونة اللغوية بالفئات النحوية الوظيفية بعد إضافة قائمة السمات داخل نظام الآلية؛ حتى يمكن للمعنون اختيار السمة التي تناسب كل كلمة، كما تظهر بالشكل رقم ٤٥.



الشكل: ٤٥ أداة Annotald أثناء العنونة بالفئات النحوية الوظيفية

- **المسند إليه SBJ**

تستخدم هذه العلامة في عدة مواضع:

- **فاعل/ نائب فاعل للفعل**

"وتصل قيمة كل جزء ١٠٠ جنيه."

```
(ROOT
 (S (NullPrefix+Conj+NullSuffix_CON و)
   (VP (Present+Active+Verb+NullSuffix_VRB تصل)
     (NP (NullPrefix+Noun+Femin+Single_SBJ قيمة)
       (NP
         (NP (NullPrefix+Noun+NullSuffix_IDF كل)
           (NP (NullPrefix+Noun+NullSuffix_IDF جزء)))
         (NP (CD_OBJ 100)
           (NP (NullPrefix+Noun+NullSuffix_TMZ جنيه))))))
   (PUNC_MRK .)))
```

- **مبتدأ، اسم كان، اسم إن، اسم كاد.**

"ودياسطي وبسنت لعب في منتهى الجدية."

```
(ROOT
 (S
   (NP (Translit_SBJ دياسطي) (NullPrefix+Conj+NullSuffix_CON و)
(Translit_NN بسنت))
   (NP (NullPrefix+Noun+NullSuffix_PRD لعب)
     (PP (NullPrefix+Prepos+NullSuffix_CON في)
       (NP (NullPrefix+Noun+ObjNoun+NullSuffix_IOB منتهى)
         (NP (Definit+Noun+RelAdj+Femin+Single_IDF الجدية)))))
   (PUNC_MRK .)))
```

- **مصدر مؤول.**

"فالأخوة يجب أن تحمل مضمون التقارب."

```
(ROOT
  (S (NullPrefix+Conj+NullSuffix_CON ف)
    (NP (Definit+Noun+NounInfinit+Femin+Single_TPC الأخوة))
    (VP (Present+Active+Verb+NullSuffix_VRB يجب)
      (SBAR (NullPrefix+ParticleNAASSIB+NullSuffix_SBJ أن)
        (S
          (VP (Present+Active+Verb+NullSuffix_VRB تحمل)
            (NP (NullPrefix+Noun+ObjNoun+NullSuffix_IOB مضمون)
              (NP (Definit+Noun+NounInfinit+NullSuffix_IDF التقارب)))))))
    (PUNC_MRK .)))
```

• المسند **PRD**

تستخدم هذه العلامة في عدة مواضع:

• **خبر لمبتدأ.**

"وهذا سيف ذو حدين."

```
(ROOT
  (S (NullPrefix+Conj+NullSuffix_CON و)
    (NP (NullPrefix+Noun+DemoPro+Masc+Single+NullSuffix_SBJ هذا))
    (NP
      (NP (NullPrefix+Noun+NullSuffix_PRD سيف))
      (NP (NullPrefix+Noun+MARF+Masc+Single+Adjunct+NullSuffix_MOD ذو)
        (NP (NullPrefix+Noun+Binary+NonAdjunct+MANSS+MAGR_IDF حدين))))
    (PUNC_MRK .)))
```

• **خبر إن وأخواتها، خبر كان وأخواتها في حالة كونه اسمًا.**

"وقال: إنه أمر آن الأوان لافتتاحه."

```
(ROOT
  (S (NullPrefix+Conj+NullSuffix_CON و)
    (VP (NullPrefix+Active+Verb+Past+NullSuffix_VRB قال) (PUNC_MRK :)
      (SBAR (NullPrefix+ParticleNAASIKH+NullSuffix_CON إن)
        (S
          (NP
            (NP (NullPrefix+Noun+Pronoun+NullSuffix_SBJ ـه))
            (NP (NullPrefix+Noun+NullSuffix_PRD أمر)))
          (VP (NullPrefix+Active+Verb+Past+NullSuffix_VRB آن)
            (NP (Definit+Noun+NounInfinit+NullSuffix_SBJ الأوان))
            (PP (NullPrefix+Prepos+NullSuffix_CON ل)
              (NP (NullPrefix+Noun+NounInfinit+NullSuffix_IOB افتتاح)
                (NP (NullPrefix+Noun+Pronoun+NullSuffix_IDF ـه)))))))))
    (PUNC_MRK .)))
```

• الموضوع TPC

تستخدم هذه العلامة في حالة:

• **إذا كان المبتدأ خبره جملة فعلية.**

"والأشجار تزدهر."

```
(ROOT
  (S (NullPrefix+Conj+NullSuffix_CON و)
    (NP (Definit+Noun+Plural+NullSuffix_TPC الأشجار))
    (VP (Present+Active+Verb+NullSuffix_VRB تزدهر))
    (PUNC_MRK .)))
```

- **إذا كان اسم كان، اسم إن، اسم كاد خبرها جملة فعلية.**

"وشدد التوضيح على أن الرئيس لم يعلن الطوارئ."

```
(ROOT
  (S (NullPrefix+Conj+NullSuffix_CON و)
    (VP (NullPrefix+Verb+NullSuffix_VRB شدد)
      (NP (Definit+Noun+NounInfinit+NullSuffix_SBJ التوضيح))
      (PP (NullPrefix+Prepos+NullSuffix_CON على)
        (SBAR (NullPrefix+ParticleNAASIKH+NullSuffix_CON أن)
          (S
            (NP (Definit+Noun+ExaggAdj+NullSuffix_TPC الرئيس))
            (VP
              (PRT (NullPrefix+JAAZIMA+NullSuffix_CON لم))
              (Present+Active+Verb+NullSuffix_VRB يعلن)
              (NP
               (NP (Definit+Noun+NoSARF+Plural+NullSuffix_OBj الطوارئ))))))))
    (PUNC_MRK .)))
```

- المفعولية OBJ

تستخدم هذه العلامة؛ للدلالة على المفعول به:

"ومارست حق الانتخاب في هيئات عدة."

```
(ROOT
  (S (NullPrefix+Conj+NullSuffix_CON و)
    (VP (NullPrefix+Verb+Past+Femin+Single_VRB مارست)
      (NP (NullPrefix+Noun+NounInfinit+NullSuffix_OBj حق)
        (NP (Definit+Noun+NounInfinit+NullSuffix_IDF الانتخاب))))
      (PP (NullPrefix+Prepos+NullSuffix_CON في)
        (NP (NullPrefix+Noun+Femin+Plural_IOB هيئات) (NullPrefix+Noun+No
unInfinit+Femin+Single_MOD عدة))))
    (PUNC_MRK .)))
```

"فالبرغم من كونه تجاوز مرحلة الصبا والشباب، إلى مرحلة كونه رجلا ناضجا، فإنه ليس صلبا عنيدا مقاوما للجديد."

```
(ROOT
  (S (NullPrefix+Conj+NullSuffix_CON ف)
    (PP (NullPrefix+Prepos+NullSuffix_CON ب)
      (NP
        (NP (Definit+Noun+NullSuffix_IOB الرغم))
        (PP (NullPrefix+Prepos+NullSuffix_CON من)
          (S
            (VP (NullPrefix+Noun+NullSuffix_IOB كون)
              (NP (NullPrefix+Noun+Pronoun+NullSuffix_IDF هـ))
              (NP (NullPrefix+Verb+NullSuffix_VRB تجاوز)
                (NP (NullPrefix+Noun+Femin+Single_OBJ مرحلة)
                  (NP (Definit+Noun+NullSuffix_IDF الصبا)
(NullPrefix+Conj+NullSuffix_CON و) (Definit+Noun+NounInfinit+NullSuffix_NN الشباب))))
              (PUNC_MRK ,)
              (PP (NullPrefix+Prepos+NullSuffix_CON إلى)
                (NP (NullPrefix+Noun+Femin+Single_IOB مرحلة)
                  (S
                    (VP (NullPrefix+Noun+NullSuffix_IDF كون)
                      (NP (NullPrefix+Noun+Pronoun+NullSuffix_IDF هـ))
                      (NP (NullPrefix+Noun+Masc+NullSuffix_OBJ رجلا) (Nul
lPrefix+Noun+SubjNoun+NullSuffix_MOD ناضجا))))))))))
    (PUNC_MRK ,)
    (PRT (NullPrefix+Conj+NullSuffix_CON ف))
    (VP (NullPrefix+ParticleNAASIKH+NullSuffix_VRB إن)
      (NP (NullPrefix+Noun+Pronoun+NullSuffix_SBj هـ))
      (S
        (VP (NullPrefix+VerbNAASIKH+NullSuffix_VRB ليس)
          (NP (NullPrefix+Noun+NullSuffix_PRD صلبا) (NullPrefix+Noun+Exag
gAdj+NullSuffix_MOD عنيدا) (NullPrefix+Noun+SubjNoun+NullSuffix_MOD مقاوما))
          (PP (NullPrefix+Prepos+NullSuffix_CON ل)
            (NP (Definit+Noun+ExaggAdj+NullSuffix_IOB الجديد))))))
    (PUNC_MRK .)))
```

- فعل VRB

تستخدم هذه العلامة؛ للدلالة على الفعل بأنواعه المختلفة.

"ويتنافس المرشحون على ٣٨٩ مقعدا."

```
(ROOT
  (S (NullPrefix+Conj+NullSuffix_CON و)
    (VP (Present+Active+Verb+NullSuffix_VRB يتنافس)
     (NP (Definit+Noun+ObjNoun+Plural+MARF+Masc+NonAdjunct_SBJ المرشحون))
      (PP (NullPrefix+Prepos+NullSuffix_CON على)
        (NP (CD_IOB 389)
          (NP (NullPrefix+Noun+NullSuffix_TMZ مقعدا)))))
    (PUNC_MRK .)))
```

- تمييز TMZ

تستخدم هذه العلامة؛ للدلالة على التمييز بأنواعه المختلفة.

"وفشل ١٣ حزبا وحركة أخرى في الحصول على أي مقعد خلال الانتخابات التشريعية الأخيرة التي جرت عام ٢٠٠٢."

```
(ROOT
  (S (NullPrefix+Conj+NullSuffix_CON و)
    (VP (NullPrefix+Noun+NullSuffix_VRB فشل)
      (NP
        (NP (CD_SBJ 13)
          (NP (NullPrefix+Noun+NullSuffix_TMZ حزبا)))
        (NullPrefix+Conj+NullSuffix_CON و)
        (NP (NullPrefix+Noun+Femin+Single_NN حركة) (NullPrefix+Noun+Femi
n+NullSuffix_MOD أخرى)))
      (PP (NullPrefix+Prepos+NullSuffix_CON في)
        (NP
          (NP (Definit+Noun+NullSuffix_IOB الحصول))
          (PP (NullPrefix+Prepos+NullSuffix_CON على)
            (NP (NullPrefix+Noun+Adjunct+NullSuffix_IOB أي)
              (NP (NullPrefix+Noun+NullSuffix_IDF مقعد))))))
      (NP (NullPrefix+Noun+NullSuffix_TMP خلال)
        (NP
          (NP (Definit+Noun+NounInfinit+Femin+Plural_IDF الانتخابات) (Defi
nit+Noun+NounInfinit+RelAdj+Femin+Single_MOD التشريعية) (Definit+Noun+Ex
aggAdj+Femin+Single_MOD الأخيرة))
          (SBAR
            (WHNP (Definit+Noun+RelPro+Femin+Single_CON التي))
            (S
              (VP (NullPrefix+Active+Verb+Past+Femin+Single_VRB جرت)
                (NP (NullPrefix+Noun+NullSuffix_OBJ عام)
                  (NP (CD_IDF 2002)))))))))
    (PUNC_MRK .)))
```

"وهنا يكون التحدي بالنسبة للمثل أكبر وأكثر تعقيدًا."

```
(ROOT
  (S (NullPrefix+Conj+NullSuffix_CON و)
    (ADVP (NullPrefix+DZARF+NullSuffix_TMP هنا))
    (VP (Present+Active+Verb+NullSuffix_VRB يكون)
      (NP (Definit+Noun+NounInfinit+NullSuffix_SBJ التحدي))
      (PP (NullPrefix+Prepos+NullSuffix_CON ب)
        (NP
          (NP (Definit+Noun+Femin+Single_IOB النسبة))
          (PP (NullPrefix+Prepos+NullSuffix_CON ل)
            (NP (Definit+Noun+SubjNoun+NullSuffix_IOB الممثل)))))
      (NP
        (ADJP (NullPrefix+Noun+NoSARF+NullSuffix_PRD أكبر)
(NullPrefix+Conj+NullSuffix_CON و) (NullPrefix+Noun+NoSARF+NullSuffix_NN أكثر)
          (NP (NullPrefix+Noun+NounInfinit+NullSuffix_TMZ تعقيدا)))))
    (PUNC_MRK .)))
```

• وصف MOD

تستخدم هذه العلامة في عدة مواضع:

• **النعت سواء أكان حقيقيًّا أو سببيًّا.**

"واحتل التجمع الوطني الديمقراطي المرتبة الثانية."

```
(ROOT
 (S (NullPrefix+Conj+NullSuffix_CON و)
   (VP (NullPrefix+Verb+NullSuffix_VRB احتل)
     (NP (Definit+Noun+NounInfinit+NullSuffix_SBJ التجمع)
(Definit+Noun+RelAdj_MOD الوطني) (Translit_MOD الديمقراطي))
     (NP (Definit+Noun+Femin+Single_OBJ المرتبة) (Definit+Noun+SubjNoun
+Femin+Single_MOD الثانية)))
   (PUNC_MRK .)))
```

• **الحال بأنواعه المختلفة.**

"وصلت جهاد إلى العمل مبكرا"

```
(ROOT
 (S (NullPrefix+Conj+NullSuffix_CON و)
   (VP (NullPrefix+Verb+Past+Femin+Single_VRB وصلت)
     (NP (NullPrefix+Noun+NounInfinit+NullSuffix_OBj جهاد))
       (PP (NullPrefix+Prepos+NullSuffix_CON إلى)
       (NP (NullPrefix+Noun+Femin+Plural_IOB العمل)))
     (NP (NullPrefix+Noun+NounInfinit+Femin+Single_MOD مبكرا)))
   (PUNC_MRK .)))
```

• إضافة IDF

تستخدم هذه العلامة؛ للدلالة على المضاف إليه.

"وخرجت قوى السلطة لتشكل المعارضة الجديدة."

```
(ROOT
 (S (NullPrefix+Conj+NullSuffix_CON و)
   (VP (NullPrefix+Active+Verb+Past+Femin+Single_VRB خرجت)
     (NP (NullPrefix+Noun+Plural+NullSuffix_SBJ قوى)
       (NP (Definit+Noun+Femin+Single_IDF السلطة)))
     (VP (ParticleNAASSIB+Present+Active+Verb+NullSuffix_VRB لتشكل)
       (NP (Definit+Noun+Femin+Single_OBJ المعارضة) (Definit+Noun+Exagg
Adj+Femin+Single_MOD الجديدة))))
   (PUNC_MRK .)))
```

• جار ومجرور IOB

تستخدم هذه العلامة؛ للدلالة على الاسم المجرور.

"فقد كانت وزيرة الخارجية كونديليزا رايس في منتهى الوضوح حين نظرت إلى حرب لبنان على أنها اللحظة التي نشهد فيها المخاض العسير لولادة الشرق الأوسط الكبير."

```
(ROOT
  (S (NullPrefix+Conj+NullSuffix_CON ف)
    (VP
      (PRT (NullPrefix+NoSyntaEffect+NullSuffix_CON قد))
      (NullPrefix+Active+Verb+Past+Femin+Single_VRB كانت)
      (NP
        (NP (NullPrefix+Noun+ExaggAdj+Femin+Single_SBJ وزيرة)
          (NP (Definit+Noun+SubjNoun+RelAdj+Femin+Single_IDF
الخارجية)))
        (NP (Translit_IDH كونديليزا) (Translit_NN رايس)))
      (PP (NullPrefix+Prepos+NullSuffix_CON في)
        (NP (NullPrefix+Noun+ObjNoun+NullSuffix_IOB منتهى)
          (NP (Definit+Noun+NullSuffix_IDF الوضوح))))
      (SBAR
        (NP (NullPrefix+DZARF+NullSuffix_TMP حين))
        (S
          (VP (NullPrefix+Active+Verb+Past+Femin+Single_VRB نظرت)
            (PP (NullPrefix+Prepos+NullSuffix_CON إلى)
              (NP (NullPrefix+Noun+NullSuffix_IOB حرب)
              (NP (NullPrefix+Noun+NoSARF+Femin+Single+NullSuffix_IDF لُبْنَان))))
            (PP (NullPrefix+Prepos+NullSuffix_CON على)
              (SBAR (NullPrefix+ParticleNAASIKH+NullSuffix_VRB أن)
                (S
                  (NP (NullPrefix+Noun+Pronoun+NullSuffix_SBJ ها))
                  (NP
                    (NP (Definit+Noun+Femin+Single_PRD اللحظة))
                    (SBAR
                     (WHNP (Definit+Noun+RelPro+Femin+Single_CON التي))
                      (S
                        (VP (Present+Active+Verb+NullSuffix_VRB نشهد)
                          (NP
                            (NullPrefix+PrepPronComp+NullSuffix_TMP
فيها)))
                        (NP
                          (NP
                            (NP (Definit+Noun+NounInfinit+NullSuffix_OBJ
المخاض) (Definit+Noun+ExaggAdj+NullSuffix_MOD العسير))
                            (PP (NullPrefix+Prepos+NullSuffix_CON ل)
                              (NP (NullPrefix+Noun+Femin+Single_IOB
ولادة)
                                (NP (Definit+Noun+NullSuffix_IDF الشرق)
(Definit+Noun+NoSARF+NullSuffix_MOD الأوسط))))))
                          (ADJP (Definit+Noun+ExaggAdj+NullSuffix_MOD
الكبير)))))))))))))
  (PUNC_MRK .))
```

- التأكيد EMP
- التوكيد اللفظي
- المفعول المطلق المؤكد

"ذلك القرار الذي تخطى خطوة تخطتها المرأة في السابق في هذا الصدد من خلال إنشاء فنادق للنساء فقط!."

```
(ROOT
  (S
    (NP
      (NP
        (NP (NullPrefix+Noun+DemoPro+Masc+Single+NullSuffix_SBJ ذلك))
        (NP (Definit+Noun+NounInfinit+NullSuffix_IDH القرار)))
      (SBAR
        (WHNP (Definit+Noun+RelPro+Masc+Single+NullSuffix_MOD الذي))
        (S
          (VP (NullPrefix+Active+Verb+Past+NullSuffix_VRB تخطى)
            (NP (NullPrefix+NounNounInfinit+Femin+Single_EMP خطوة))
            (S
              (VP (NullPrefix+Verb+Past+Femin+Single_VRB خطت)
                (NP (NullPrefix+Noun+Pronoun+NullSuffix_OBJ ها))
                (NP (Definit+Noun+Femin+Single_SBJ المرأة))
                (PP (NullPrefix+Prepos+NullSuffix_CON في)
                  (NP (Definit+Noun+SubjNoun+NullSuffix_IOB السابق)))
                (PP (NullPrefix+Prepos+NullSuffix_CON في)
                  (NP
                    (NP
                      (NP (NullPrefix+Noun+DemoPro+Masc+Single+NullSuff
ix_IOB هذا))
                      (NP (Definit+Noun+NullSuffix_IDH الصدد)))
                    (PP (NullPrefix+Prepos+NullSuffix_CON من)
                      (NP (NullPrefix+Noun+NullSuffix_TMP خلال)
                       (NP (NullPrefix+Noun+NounInfinit+NullSuffix_IDF إنشاء)
                           (NP
                             (NP (NullPrefix+Noun+NoSARF+Plural+NullSuff
ix_IDF فنادق))
                             (PP (NullPrefix+Prepos+NullSuffix_CON ل)
                               (NP (Definit+Noun+Plural+NullSuffix_IOB
النساء)))))))))
                (ADVP (NullPrefix+Noun+NullSuffix_MOD فقط)))))))
    (PUNC_MRK !) (PUNC_MRK .)))
```

- الزمان والمكان TMB

تستخدم هذه العلامة للدلالة على المفعول فيه (ظرف الزمان والمكان).

"وفشل ١٣ حزبا وحركة أخرى في الحصول على أي مقعد خلال الانتخابات التشريعية الأخيرة التي جرت عام ٢٠٠٢."

```
(ROOT
  (S (NullPrefix+Conj+NullSuffix_CON و)
    (VP (NullPrefix+Noun+NullSuffix_VRB فشل)
      (NP
        (NP (CD_SBJ 13)
          (NP (NullPrefix+Noun+NullSuffix_TMZ حزبا)))
        (NullPrefix+Conj+NullSuffix_CON و)
        (NP (NullPrefix+Noun+Femin+Single_NN حركة) (NullPrefix+Noun+Femi
n+NullSuffix_MOD أخرى)))
      (PP (NullPrefix+Prepos+NullSuffix_CON في)
        (NP
          (NP (Definit+Noun+NullSuffix_IOB الحصول))
          (PP (NullPrefix+Prepos+NullSuffix_CON على)
            (NP (NullPrefix+Noun+Adjunct+NullSuffix_IOB أي)
              (NP (NullPrefix+Noun+NullSuffix_IDF مقعد))))))
      (NP (NullPrefix+Noun+NullSuffix_TMP خلال)
        (NP
          (NP (Definit+Noun+NounInfinit+Femin+Plural_IDF الانتخابات) (Defi
nit+Noun+NounInfinit+RelAdj+Femin+Single_MOD التشريعية) (Definit+Noun+Ex
aggAdj+Femin+Single_MOD الأخيرة))
          (SBAR
            (WHNP (Definit+Noun+RelPro+Femin+Single_CON التي))
            (S
              (VP (NullPrefix+Active+Verb+Past+Femin+Single_VRB جرت)
                (NP (NullPrefix+Noun+NullSuffix_OBJ عام)
                  (NP (CD_IDF 2002)))))))))
    (PUNC_MRK .)))
```

- السببية PRB

تستخدم هذه العلامة؛ للدلالة على المفعول لأجله.

"وأمي تمنعني من اللعب بها في الشارع؛ خوفا من السيارات"

```
(ROOT
 (S (NullPrefix+Conj+NullSuffix_CON و)
  (NP (NullPrefix+Noun+NullSuffix_TPC أم)
    (NP (NullPrefix+Noun+Pronoun+NullSuffix_IDF ي)))
  (VP (Present+Active+Verb+NullSuffix_VRB تمنع)
    (NP (NullPrefix+Noun+Pronoun+NullSuffix_OBJ ي))
    (PP (NullPrefix+Prepos+NullSuffix_CON من)
      (NP (Definit+Noun+NullSuffix_IOB اللعب))
        (NP (NullPrefix+PrepPronComp+NullSuffix_IOB بها))))
  (PP (NullPrefix+Prepos+NullSuffix_CON في)
    (NP (Definit+Noun+SubjNoun+NullSuffix_IOB الشارع)))
  (PUNC_MRK ;)
  (NP
    (NP (NullPrefix+Noun+NullSuffix_PRB خوفا))
    (PP (NullPrefix+Prepos+NullSuffix_CON من)
      (NP (Definit+Noun+ExaggAdj+Femin+Plural_IOB السيارات))))))
```

- الإيضاح IDH

تستخدم هذه العلامة في عدة مواضع:

- **عطف البيان**

"قال المحلل السياسي اللبناني المستقل سركيس أبو زيد علم."

```
(ROOT
  (S
    (VP (NullPrefix+Active+Verb+Past+NullSuffix_VRB قال)
      (NP
        (NP (Definit+Noun+SubjNoun+NullSuffix_SBj المحلل)
(Definit+Noun+RelAdj_MOD السياسي) (Definit+Noun+NoSARF+Femin+Single+RelA
dj_MOD اللبناني) (Definit+Noun+SubjNoun+NullSuffix_MOD المستقل))
        (NP (Translit_IDH سركيس) (NullPrefix+Noun+NullSuffix_NN أبوزيد)
(NullPrefix+Noun+NullSuffix_NN علم))))
    (PUNC_MRK .)))
```

- **المفعول المطلق سواء أكان مبينا للنوع أو للعدد.**

"أساءت إلى جنس النساء إساءة كبرى."

```
(ROOT
  (S
    (NP (NullPrefix+Verb+Past+Femin+Single_VRB أساءت)
      (NP
        (NP (NullPrefix+Prepos+NullSuffix_CON إلى))
        (NP (NullPrefix+Noun+NullSuffix_IOB جنس)
          (NP (Definit+Noun+Plural+NullSuffix_IDF النساء)))))
    (NP (NullPrefix+Noun+NounInfinit+Femin+Single_IDH إساءة) (NullPrefix+
Noun+Femin+NullSuffix_MOD كبرى))
    (PUNC_MRK .)))
```

- **الاختصاص**

"ونحن الحزب الوحيد الذي ترشح وهو ينتمي للأغلبية."

```
 (ROOT
  (S (NullPrefix+Conj+NullSuffix_CON و)
    (NP (NullPrefix+Noun+SubjPro+NullSuffix_SBJ نحن))
    (NP
      (NP (Definit+Noun+NullSuffix_IDH الحزب) (Definit+Noun+ExaggAdj+Null
Suffix_MOD الوحيد))
        (SBAR
          (WHNP (Definit+Noun+RelPro+Masc+Single+NullSuffix_MOD الذي))
          (S
            (VP (NullPrefix+Verb+NullSuffix_VRB ترشح)
              (SBAR (NullPrefix+Conj+NullSuffix_CON و)
                (S
                  (NP (NullPrefix+Noun+SubjPro+NullSuffix_TPC هو))
                  (VP (Present+Active+Verb+NullSuffix_VRB ينتمي)
                    (PP (NullPrefix+Prepos+NullSuffix_CON إلى)
                      (NP (Definit+Noun+NoSARF+RelAdj+Femin+Single_IOB
الأغلبية))))))))))
    (PUNC_MRK .)))
```

- **البدل بأنواعه.**

"ولم يعاقبه الاتحاد الدولي لكرة القدم الفيفا."

```
(ROOT
  (S (NullPrefix+Conj+NullSuffix_CON و)
    (VP
      (PRT (NullPrefix+JAAZIMA+NullSuffix_CON لـم))
      (Present+Active+Verb+NullSuffix_VRB يـعـاقـب)
      (NP (NullPrefix+Noun+Pronoun+NullSuffix_OBJ ـه))
      (NP
        (NP
          (NP (Definit+Noun+NounInfinit+NullSuffix_SBJ الاتـحـاد)
(Definit+Noun+Plural+RelAdj_MOD الـدولـي))
          (PP (NullPrefix+Prepos+NullSuffix_CON ل)
            (NP (NullPrefix+Noun+Femin+Single_IOB كـرة)
              (NP (Definit+Noun+NullSuffix_IDF الـقـدم)))))
        (NP (Translit_IDH الـفـيـفـا))))
    (PUNC_MRK .)))
```

• المتلازم NN

تستخدم هذه العلامة؛ للدلالة على أحد أجزاء التلازم التركيبي (البنية المسطحة)[1] الذي يكون في حكم المفرد نحويًّا، والذي يتمثل في بنية مسطحة وفق التنظيم البنائي للمركبات.

• **أسماء الأعلام المركبة تركيبًا إسناديًّا.**

"أكد الناقد السينمائي السعودي خالد ربيع السيد، أن مهرجان جدة الثالث للأفلام، أخذ الأهمية الكبرى عن التجربتين السابقتين."

```
(ROOT
  (S
    (VP (NullPrefix+Verb+NullSuffix_VRB أكـد)
      (NP
        (NP
          (NP (Definit+Noun+SubjNoun+NullSuffix_SBJ الـنـاقـد) (Translit_
MOD الـسـيـنـمـائـي))
          (ADJP (Definit+Noun+RelAdj_MOD الـسـعـودي)))
        (PUNC_MRK ,)
        (NP (NullPrefix+Noun+SubjNoun+NullSuffix_IDH خـالـد)
            (NullPrefix+Noun+ExaggAdj+NullSuffix_NN ربـيـع)
            (Definit+Noun+ExaggAdj+NullSuffix_NN الـسـيـد)))
      (PUNC_MRK ,)
      (SBAR (NullPrefix+ParticleNAASIKH+NullSuffix_VRB أن)
        (S
          (NP
            (NP (NP (NullPrefix+Noun+NullSuffix_TPC مـهـرجـان)
               (NP (NullPrefix+Noun+Femin+Single_IDF جـدة)))  (ADJP (Defi
nit+Noun+SubjNoun+NullSuffix_MOD الـثـالـث)))
            (PP (NullPrefix+Prepos+NullSuffix_CON ل)
              (NP (Definit+Noun+NullSuffix_IOB الأفـلام))))
          (PUNC_MRK ,)
          (VP (NullPrefix+Active+Verb+Past+NullSuffix_VRB أخـذ)
            (NP (Definit+Noun+NoSARF+RelAdj+Femin+Single_SBJ الأهـمـيـة)
(Definit+Noun+Femin+NullSuffix_MOD الـكـبـرى))
            (PP (NullPrefix+Prepos+NullSuffix_CON عن)
              (NP (Definit+Noun+NounInfinit+Femin+Binary+NonAdjunct+MANS
S+MAGR_IOB الـتـجـربـتـيـن) (Definit+Noun+SubjNoun+Femin+Binary+NonAdjunct+MA
NSS+MAGR_MOD الـسـابـقـتـيـن)))))))
    (PUNC_MRK .)))
```

(١) يقصد بالبنية المسطحة البنية التي توازي بين أحد أجزائها، بحيث تكتسب كل الأجزاء صفة الابن لرأس المركب.

● **الأسماء المبنية على فتح الجزأين.**

"مر واحد وثلاثون عاما على التحاقي بالعمل الاجتماعي، حيث بدأته وأنا ابنة السابعة عشرة."

```
 (ROOT
  (S
    (VP (NullPrefix+Active+Verb+Past+NullSuffix_VRB مـر)
      (NP (NullPrefix+Noun+SubjNoun+NullSuffix_SBJ واحـد)
(NullPrefix+Conj+NullSuffix_CON و) (NullPrefix+Noun+Plural+MARF+Masc+NonA
djunct_NN ثـلاثـون)
        (NP (NullPrefix+Noun+NullSuffix_TMZ عـامـا)))
      (PP (NullPrefix+Prepos+NullSuffix_CON عـلـى)
        (NP (NullPrefix+Noun+NounInfinit+NullSuffix_IOB الـتـحـاق)
          (NP (NullPrefix+Noun+Pronoun+NullSuffix_IDF ي))
          (PP (NullPrefix+Prepos+NullSuffix_CON ب)
            (NP (Definit+NounNounInfinit+NullSuffix_IOB الـعـمـل)
(Definit+Noun+RelAdj_MOD الاجـتـمـاعـي)))))
      (PUNC_MRK ,)
      (SBAR
        (WHADVP (NullPrefix+DZARF+NullSuffix_TMP حـيـث))
        (S
          (VP (NullPrefix+Active+Verb+Past+SubjPro_VRD بـدأت)
            (NP (NullPrefix+Noun+Pronoun+NullSuffix_OBJ ـهـ))
            (NullPrefix+Conj+NullSuffix_CON و)
            (S
              (NP (NullPrefix+Noun+SubjPro+NullSuffix_SBJ أنـا))
              (NP (NullPrefix+Noun+Femin+Single_PRD ابـنـة)
                (NP (Definit+Noun+SubjNoun+Femin+Single_IDF الـسـابـعـة)
(NullPrefix+Noun+Femin+Single_NN عشرة))))))))
    (PUNC_MRK .)))
```

● **العطف**

"دياسطي وبسنت لعب في منتهى الجدية."

```
(ROOT
  (S
    (NP (Translit_SBJ ديـاسـطـي) (NullPrefix+Conj+NullSuffix_CON و)
(Translit_NN بـسـنـت))
    (NP (NullPrefix+Noun+NullSuffix_PRD لـعـب)
      (PP (NullPrefix+Prepos+NullSuffix_CON فـي)
        (NP (NullPrefix+Noun+ObjNoun+NullSuffix مـنـتـهـى)
          (NP (Definit+Noun+RelAdj+Femin+Single الـجـديـة)))))
    (PUNC_MRK .)))
```

● فعل_فاعل VRD

تستخدم هذه العلامة؛ للدلالة على الأفعال المسندة للضمائر المتصلة المبنية في محل رفع فاعل.

"واستمر الزعماء التقليديون ممن أداروا الحرب الأهلية ممسكين بزمام الأمور."

```
(ROOT
  (S (NullPrefix+Conj+NullSuffix_CON و)
    (VP (NullPrefix+Verb+NullSuffix_VRB استمر)
      (NP
        (NP (Definit+Noun+NoSARF+Plural+NullSuffix_SBJ الزعماء) (Definit+
Noun+NounInfinit+Plural+MARF+RelAdj+Masc+NonAdjunct_MOD التقليديون))
        (SBAR
          (WHNP (NullPrefix+Prepos+Noun+RelPro+NullSuffix_IOB ممن))
          (S
            (VP (NullPrefix+Active+Verb+Past+SubjPro_VRD أداروا)
              (NP (Definit+Noun+NullSuffix_OBJ الحرب) (Definit+Noun+RelAd
j+Femin+Single_MOD الأهلية))
              (S
                (VP (NullPrefix+Noun+SubjNoun+Plural+Masc+NonAdjunct+MA
NSS+MAGR_MOD ممسكين)
                  (PP (NullPrefix+Prepos+NullSuffix_CON ب)
                    (NP (NullPrefix+Noun+NullSuffix_IOB زمام)
                      (NP (Definit+Noun+NullSuffix_IDF الأمور))))))))))))
    (PUNC_MRK .)))
```

- ربط CON

تستخدم هذه العلامة لجميع أدوات الربط:

- **أدوات العطف**

"وتحولت المحكمة إلى ورقة بيد الصراعات الدولية."

```
(ROOT
  (S (NullPrefix+Conj+NullSuffix_CON و)
    (VP (NullPrefix+Verb+Past+Femin+Single_VRB تحولت)
      (NP (Definit+Noun+Femin+Single_SBJ المحكمة))
      (PP (NullPrefix+Prepos+NullSuffix_CON إلى)
        (NP
          (NP (NullPrefix+Noun+Femin+Single_IOB ورقة))
          (PP (NullPrefix+Prepos+NullSuffix_CON ب)
            (NP (NullPrefix+Noun+Femin+NullSuffix_IOB يد)
              (NP (Definit+Noun+Femin+Plural_IDF الصراعات) (Definit+Noun
+RelAdj+Femin+Single_MOD الدولية)))))))
    (PUNC_MRK .)))
```

- **أدوات نصب الفعل المضارع**

"لن تصطدم معهم أبدا."

```
(ROOT
  (S
    (VP
      (PRT (NullPrefix+ParticleNAASSIB+NullSuffix_CON لن))
      (Present+Active+Verb+NullSuffix_VRB نصطدم)
      (NP (NullPrefix+DZARF+Adjunct+NullSuffix_TMP مع)
        (NP (NullPrefix+Noun+Pronoun+NullSuffix_IDF هم)))
      (NP (NullPrefix+Noun+NullSuffix_TMP أبدا)))
    (PUNC_MRK .)))
```

• واو الحال

"ونحن الحزب الوحيد الذي ترشح وهو ينتمي للأغلبية."

```
(ROOT
  (S (NullPrefix+Conj+NullSuffix_CON و)
    (NP (NullPrefix+Noun+SubjPro+NullSuffix_SBJ نحن))
    (NP
      (NP (Definit+Noun+NullSuffix_IDH الحزب) (Definit+Noun+ExaggAdj+Null
Suffix_MOD الوحيد))
        (SBAR
          (WHNP (Definit+Noun+RelPro+Masc+Single+NullSuffix_MOD الذي))
          (S
            (VP (NullPrefix+Verb+NullSuffix_VRB ترشح)
              (SBAR (NullPrefix+Conj+NullSuffix_CON و)
                (S
                  (NP (NullPrefix+Noun+SubjPro+NullSuffix_TPC هو))
                  (VP (Present+Active+Verb+NullSuffix_VRB ينتمي)
                    (PP (NullPrefix+Prepos+NullSuffix_CON إلى)
                      (NP (Definit+Noun+NoSARF+RelAdj+Femin+Single_IOB
الأغلبية))))))))))
    (PUNC_MRK .)))
```

• واو المفعولية

"وشدد التوضيح على أن الرئيس لم يعلن الطوارئ والمساء."

```
(ROOT
  (S (NullPrefix+Conj+NullSuffix_CON و)
    (VP (NullPrefix+Verb+NullSuffix_VRB شدد)
      (NP (Definit+Noun+NounInfinit+NullSuffix_SBJ التوضيح))
      (PP (NullPrefix+Prepos+NullSuffix_CON على)
        (SBAR (NullPrefix+ParticleNAASIKH+NullSuffix_VRB أن)
          (S
            (NP (Definit+Noun+ExaggAdj+NullSuffix_TPC الرئيس))
            (VP
              (PRT (NullPrefix+JAAZIMA+NullSuffix_CON لم))
              (Present+Active+Verb+NullSuffix_VRB يعلن)
                (NP (Definit+Noun+NoSARF+Plural+NullSuffix_OBJ
الطوارئ)))
                (NullPrefix+Conj+NullSuffix_CON و)
                (NP (Definit+Noun+ NullSuffix_OBJ المساء)) ))))
    (PUNC_MRK .)))
```

• حروف الجر

"وهي الحالات التاريخية لعمل المرأة."

```
(ROOT
  (S (NullPrefix+Conj+NullSuffix_CON و)
    (NP (NullPrefix+Noun+SubjPro+NullSuffix_SBJ هي))
    (NP
      (NP (Definit+Noun+Femin+Plural_PRD الحالات) (Definit+Noun+RelAdj+Fe
min+Single_MOD التاريخية))
      (PP (NullPrefix+Prepos+NullSuffix_CON ل)
        (NP (NullPrefix+NounNounInfinit+NullSuffix_IOB عمل)
          (NP (Definit+Noun+Femin+Single_IDF المرأة)))))
    (PUNC_MRK .)))
```

• **أدوات الشرط**

"وإذا كانت الأخبار سلبية فستتأثر البورصة."

```
(ROOT
 (S (NullPrefix+Conj+NullSuffix_CON و)
   (SBAR (NullPrefix+CondNotJAAZIMA+NoSyntaEffect+NullSuffix_CON إذا)
     (S
       (VP (NullPrefix+Active+Verb+Past+Femin+Single_VRB كانت)
         (NP (Definit+Noun+Plural+NullSuffix_SBJ الأخبار))
         (ADJP (NullPrefix+Noun+RelAdj+Femin+Single_PRD سلبية)))))
   (PRT (NullPrefix+Conj+NullSuffix_CON ف))
   (VP (Future+Present+Active+Verb+NullSuffix_VRB ستتأثر)
     (NP (Translit_SBJ البورصة)))
   (PUNC_MRK .)))
```

• **أدوات الاستثناء**

"فليس هناك هذا العائد، إلا في أحلام السذج."

```
(ROOT
 (S (NullPrefix+Conj+NullSuffix_CON ف)
   (VP (NullPrefix+VerbNAASIKH+NullSuffix_VRB ليس)
     (ADVP (NullPrefix+DZARF+NullSuffix_PRD هناك))
     (NP
       (NP (NullPrefix+Noun+DemoPro+Masc+Single+NullSuffix_SBJ هذا))
       (NP (Definit+Noun+SubjNoun+NullSuffix_IDH العائد)))
     (PUNC_MRK ,)
     (PP
       (PRT (NullPrefix+Except+NullSuffix_CON إلا))
       (NullPrefix+Prepos+NullSuffix_CON في)
       (NP (NullPrefix+Noun+Plural+NullSuffix_IOB أحلام)
         (NP (Definit+Noun+Plural+NullSuffix_IDF السذج)))))
   (PUNC_MRK .)))
```

• علامة MRK

تستخدم هذه العلامة لجميع علامات الترقيم والرموز الرياضية المستخدمة في النص.

"ويرأس الوزارة مسلم سني ومجلس النواب مسلم شيعي."

```
(ROOT
  (S (NullPrefix+Conj+NullSuffix_CON و)
    (VP (Present+Active+Verb+NullSuffix_VRB يرأس)
      (NP
        (NP (Definit+Noun+Femin+Single_OBJ الوزارة))
        (NP (NullPrefix+Noun+SubjNoun+NullSuffix_SBJ مسلم)
(NullPrefix+Noun+RelAdj_MOD سني))
        (PUNC_MRK ,) (NullPrefix+Conj+NullSuffix_CON و)
        (NP (NullPrefix+Noun+NullSuffix_NN مجلس)
          (NP (Definit+Noun+Plural+NullSuffix_IDF النواب)))
        (NP (NullPrefix+Noun+SubjNoun+NullSuffix_SBJ مسلم)
(NullPrefix+Noun+RelAdj_MOD شيعي))))
    (PUNC_MRK .)))
```

٥. عناصر هيكلة البناء النحوي

يتم تنظيم البنية التركيبية للجملة في البناء الشجري - كما ظهرت في الأمثلة السابقة - من خلال بنيتي العوامل والوصف اللتين تفرضهما النظرية النحوية على تشكيل عناصر المركبات الأساسية لاسيما الفعلية والاسمية، وذلك في إطار القوالب التنظيمية للتمثيل المكوني.

٥-١ بنية العوامل Arguments structure

تحدد النظرية النحوية العوامل التي يتطلبها المحمول Predicate سواء أكان اسما أو فعلا.

- **الفعل أو المركب الفعلي**

تعيّن دلالة الفعل أو المركب الفعلي عددًا من العوامل مثل:

- المركب الاسمي (الفاعل SBJ)
- المركب الاسمي (المفعول OBJ)
- المركب الحرفي (المفعول غير المباشر IOB)
- الجملة التكميلية (جملة الصلة SBAR)

وتتدرج هذه العوامل في البنية الشجرية في المستوى الذي يلي المركب الفعلي أي أبناء لهذا المركب، وأخوات لرأسه؛ أي الفعل.

"فقدم الصورة الصحيحة للصوفية، حيث اكتظاظها بالعديد من البدع."

```
(ROOT
  (S (NullPrefix+Conj+NullSuffix ف)
    (VP (NullPrefix+Verb+NullSuffix قدم)
      (NP
        (NP (Definit+Noun+Femin+Single الصورة) (Definit+Noun+ExaggAdj+Fe
min+Single الصحيحة))
        (PP (NullPrefix+Prepos+NullSuffix ل)
          (NP (Definit+Noun+RelAdj+Femin+Single الصوفية))))
      (PUNC_MRK ;)
      (SBAR
        (WHADVP (NullPrefix+CondJAAZIMA+DZARF+NullSuffix حيث))
        (S
          (NP
            (NP (NullPrefix+Noun+NounInfinit+NullSuffix اكتظا)
              (NP (NullPrefix+Noun+Pronoun+NullSuffix ها)))
            (PP (NullPrefix+Prepos+NullSuffix ب)
              (NP
                (NP (Definit+Noun+ExaggAdj+NullSuffix العديد))
                (PP (NullPrefix+Prepos+NullSuffix من)
                  (NP (Definit+Noun+NullSuffix البدع)))))))))
    (PUNC_MRK .)))
```

• الاسم أو المركب الاسمي

تعيّن دلالة الاسم أو المركب الاسمي بنية العامل المتمثلة في:

• المركب الاسمي (المضاف إليه IDF)

ويتدرج هذا العامل (المكمل) في البنية الشجرية في المستوى الذي يلي المركب الاسمي أو الاسم (الرأس) أي يكون ابنا لرأس المركب الاسمي.

"سر الانتصار."

```
(ROOT
  (FRAG
    (NP (NullPrefix+Noun+NullSuffix سر)
      (NP (Definit+Noun+NounInfinit+NullSuffix الانتصار)))
    (PUNC_MRK .)))
```

٥-٢ بنية الوصف Adjunct Structure

تعددت أضرب الوصف أو التوسعة التي ترتبط بالفعل أو المركب الفعلي و الاسم أو المركب الاسمي؛ لتقوية المعنى، وتحديد الأبعاد التي يسعى إليها المتكلم.

• الفعل أو المركب الفعلي

يتوسع الفعل في تكوين علاقات تركيبية (فضلة)؛ للزيادة في المعنى، مثل:

-المركب الاسمي (التأكيد، التمييز، الإيضاح)

- المركب الحرفي

-المركب الظرفي

وتشغل هذه المركبات المستوى الذي يلي المركب الفعلي في البنية الشجرية.

"وتحاول تنوع مصادرها، بدلا من الاعتماد على مصدر واحد فقط هو البترول."

```
(ROOT
 (S (NullPrefix+Conj+NullSuffix و)
   (VP (Present+Active+Verb+NullSuffix تحاول)
     (NP (NullPrefix+Noun+NounInfinit+NullSuffix تنويع)
       (NP (NullPrefix+Noun+NoSARF+Plural+NullSuffix مصادر)
         (NP (NullPrefix+Noun+Pronoun+NullSuffix ها))))
     (PUNC_MRK ,)
     (NP
       (NP (NullPrefix+Noun+NullSuffix بدلا))
       (PP (NullPrefix+NullSuffix من)
         (NP
           (NP
             (NP
               (NP (Definit+Noun+NounInfinit+NullSuffix الاعتماد))
               (PP (NullPrefix+Prepos+NullSuffix على)
                 (NP (NullPrefix+Noun+NullSuffix مصدر) (NullPrefix+Noun+
SubjNoun+NullSuffix واحد) (NullPrefix+Noun+NullSuffix فقط))))
             (NP (NullPrefix+Noun+SubjPro+NullSuffix هو)))
           (NP (Translit البترول))))))
    (PUNC_MRK .)))
```

• الاسم أو المركب الاسمي

يتوسع الاسم بنشوء علاقة وصفية بين النعت والمنعوت؛ لإزالة الإبهام، وبيان المعنى، مثل:

-الوصف (مركب الصفة)

- المركب الحرفي (في حالة كونه وصفًا)

-الوصف الذي يصف المركب الاسمي (MOD)

وتشغل هذه المركبات المستوى الذي يلي المركب الاسمي في البنية الشجرية، موازيًة الاسم المضاف إليه (عامل الاسم) أي في بنية مسطحة Flat Structure.

"حزب التجمع الوطني الديمقراطي."

```
(ROOT
  (FRAG
    (NP (NullPrefix+Noun+NullSuffix حزب)
      (NP (Definit+Noun+NounInfinit+NullSuffix التجمع) (Definit+Noun+RelAdj
الوطني) (Translit الديمقراطي)))
    (PUNC_MRK .)))
```

•••

الفصل الرابع

ترميز البنك الشجري النحوي

يعنى ترميز البنك الشجري النحوي بتحويله من صورته الأساسية أي الأقواس التركيبية Syntactic Bracketing بما تتضمنه من معلومات نحوية إلى صورة نصية مذيلة بمعلومات رئيسية حول العناصر الأساسية المكونة للتحليل الشجري، وذلك باستخدام لغة ترميز قياسية، قادرة على ترميز العديد من البيانات المختلفة.

وتعد لغة الترميز القابلة للامتداد[1] (Extensible Markup Language (XML لغةً توصيفيةً مثاليةً، حيث تدعّم نظام الحرف الدولي الموحَّد للغات الطبيعية، فضًلا عن أن لديها القدرة على تمثيل جميع أنواع الوثائق النصية، وقواعد البيانات، ولديها قابلية التعامل مع جميع الأنظمة الحاسوبية.[2]

وتظهر أهمية ترميز المدونات اللغوية بشكل عام عند توظيفها أو استدعائها في الآليات البرمجية التي تتطلب معلومات تمييزية حول نصوص هذه المدونات، مثل: حدود الفقرة،

(١) هي في الأساس مجموعة قواعد مبسطة من لغة الترميز ذات المعيار المعمم Standard Generalized Markup (SGM) التي أنشئت في بدايات الثمانيات من القرن المنصرم، وعلى الرغم من زيادة الطلب عليها آنذاك في تخزين البيانات وتنظيمها إلا أنها كانت لديها بعض الثغرات في تنظيم نصوص صفحات الويب؛ الأمر الذي أدى إلى انبثاق لغة الترميز القابلة للامتداد منها؛ لحل هذه الثغرات. ينظر:

Birbeck, M. & Others. (2001). Professional XML 2nd Edition. Published by Wrox Press Ltd, the United States. P.5-8.

(٢) السعيد (المعتزبالله): مدونة معجم تاريخي للغة العربية، ص ٥١.

حدود الجملة، حدود الكلام، فواصل الفقرة،...إلخ، لأن النص العادي Plain Text لا يقدم هذه المعلومات البسيطة للآليات البرمجية المختلفة. فالهدف من الترميز هو وصف البيانات وتنظيمها، بحيث يمكن استدعاؤها بسهولة لدى الأنظمة الحاسوبية.

١ - لغة الترميز القابلة للامتداد XML

هي مجموعة من العناصر والوسوم، يتم نظمها وفق قواعد كتابية معينة على النص المراد ترميزه، وتتمثل هذه الوسوم بين الأقواس المثلثة < الوسم > لتعبّر عما هو خارجها أو تدّل عليه. [1] ويمكن توضيح ذلك من خلال تطبيق الوسم الذي يشير إلى حدود الجملة: <S> على المثال التالي:

”هذه ظاهرة قد يسببها الخوف. يتوقع محللون سياسيون أن يكون هادئا.“

وباستخدام XML، يمكن تحديد حدود الجملة:

<S> هذه ظاهرة قد يسببها الخوف. <S> <S> يتوقع محللون سياسيون أن يكون هادئا.<S>

ويمكن أن تضاف العديد من الوسوم المعبرة عن دلالات النص أو الرمز أو قواعد البيانات داخل وثيقة XML وفق ضوابط ترتيبية معينة من خلال نظمها في ملف تعريفي يدعى بتعريفات نوع الوثيقة Document Type Definitions (DTD) يتضمن تلك الوسوم بضوابطها التي يستوجب على ملف XML اتباعها.

ويتم ربط ملف تعريفات نوع الوثيقة المنفصل بمستندات XML من خلال كتابة مساره داخل الملف الذي يقوم بعمل تحليل Parsing للملف النصي المراد هيكلته بلغة الترميز القابلة للامتداد.

(1) Gulbransen, D. (2002).Ibid.P.9.

٢- ترميز البنك الشجري النحوي بلغة XML

تم ترميز البنك الشجري النحوي بلغة الترميز القابلة للامتداد XML، باستخدام ملف تعريفات نوع الوثيقة DTD[١] الخاص بمشروع البنك الشجري الألماني[٢] TIGER الذي يتضمن بيانات تعريفية حول تحليلات التقويس النحوي للجمل، مرتبة كالآتي:

- تعريف المدونة Corpus ID.
- الرقم الترتيبي للجملة Sentence ID.
- تحديد نقطة بداية الجملة Graph root.
- تحديد الأجزاء التي تتضمنها العقد النهائية Terminal في التقويس، والتي اشتملت على:
  - الرقم الترتيبي للكلمة في الجملة S1_1 ,S1_2, S1_3
  - الكلمة Word
  - الأقسام الكلامية POS
- تحديد الأجزاء التي تتضمنها العقد غير النهائية NoN-Terminal في التقويس، والتي اشتملت على:
  - الرقم الترتيبي للعقدة Node، والذي يبدأ من الرقم ٥٠٠ مثال ذلك S1_500, S1_502
  - الفئة التركيبية التي تقابل رقم العقدة.

ويوضح الشكل رقم ٤٦ جزءًا من ملف تعريفات نوع الوثيقة DTD التي اعتمدت عليها الدراسة في تحديد عناصر تنسيق ملف XML:

(١) الملف الذي يعني بوصف التحليل النحوي في المدونة اللغوية.

(٢) ينظر:
TIGER Corpus. (2003). http://www.ims.uni-stuttgart.de/forschung/ressourcen/korpora/tiger.en.html

```
<!-- declaration of the root element: corpus -->

<xsd:element name="corpus">

  <xsd:complexType>

    <xsd:sequence>

      <xsd:choice>
        <!-- header of the document is optional -->
        <xsd:element name="head" type="headType" minOccurs="0" maxOccurs="1"/>
      </xsd:choice>

      <xsd:element name="body" type="bodyType" minOccurs="1" maxOccurs="1"/>

    </xsd:sequence>

    <!-- corpus ID -->
    <xsd:attribute name="id"      type="idType"     use="required" />

    <!-- optional attribute: TigerXML version; used by TIGERSearch only -->
    <xsd:attribute name="version" type="xsd:string" use="optional" />

  </xsd:complexType>

</xsd:element>
```

الشكل:٤٦ نموذج من تعريفات نوع الوثيقة DTD[١]

يمثل الشكل السابق تعريفات عنصر المدونة، إذ يتضمن السمات attributes التي تشكِّل عنصر تعريف المدونة المتمثلة في الاسم= ID، ونوع الاسم = idtType من حيث كونه (سلسلة نصية أو رقمية) وإدخال المستخدم لهذا الاسم Required.

ثم استخدمت الدراسة أداة TIGERRegistry[٢]– المتضمنة في أدوات بحث تايجر TIGERSearchTools الخاصة بالبنك الشجري الألماني – باعتبارها محلّلا Parser لمستند الوثيقة بلغة XML، وذلك في إطار ملف تعريفات نوع الوثيقة DTD.

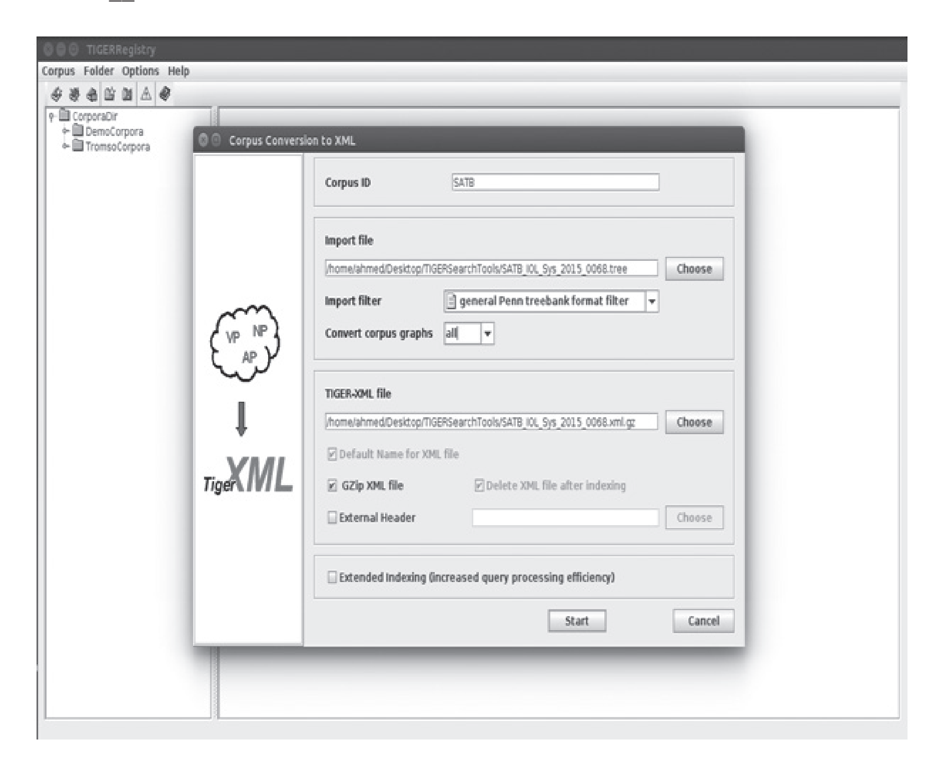
ويوضح الشكل رقم ٤٧ أداة TIGERRegistry أثناء تحويلها للملف الشجري (التقويسي) أدناه إلى ملفٍ بصيغة لغة الترميز القابلة للامتداد XML.

(١) للاطلاع على تعريفات الوثيقة:
TIGER DTD. (2003).www.ims.uni-stuttgart.de/forschung/ressourcen/korpora/tiger.en.html

(٢) هي أداة من ضمن حزمة package أدوات النظام الألماني البحثي TIGERSearchTools الذي يرعاه معهد معالجة اللغة الطبيعية بجامعة شتوتغارت University of Stuttgart، والذي يهدف إلى البحث في المدونات اللغوية المعنونة لا سيما التركيبية.

"و هو رهان شجاع واثق، مرة أخرى." إسلام أون لاين- ٢٠٠٨.

```
(S (NullPrefix+Conj+NullSuffix_CON و)
  (NP (NullPrefix+Noun+SubjPro+NullSuffix_SBJ هُوَ))
  (NP (NullPrefix+Noun+NullSuffix_PRD رِهَانٌ)
    (NP
      (ADJP (NullPrefix+Noun+NullSuffix_MOD شُجَاعٌ)
            (NullPrefix+Conj+NullSuffix_CON و)
            (NullPrefix+Noun+SubjNoun+NullSuffix_COJ وَاثِقٌ))))
  (PUNC__ ،)
  (NP
    (NP (NullPrefix+Noun+Femin+Single_MOD مَرَّةً))
    (ADJP (NullPrefix+Noun+Femin+NullSuffix_MOD أُخْرَى)))
  (PUNC__ .))
```



الشكل:٤٧ أداة **TIGERRegistry** الألمانية [(١)]

يظهر في الشاشة السابقة خطوات ترميز أو تحويل ملفات أقواس التحليل إلى ملفات مهيكلة بصيغة XML، والتي تمثلت فيما يلي:

(١) يمكن تحميله من خلال هذا الرابط:
TIGERRegistry. (2003). www.ims.uni-stuttgart.de/forschung/ressourcen/werkzeuge/tigersearch.html

- إدخال تعريف المدونة Corpus ID.
- تحديد مسار الملف التحليلي المراد تحويله.
- تحديد نوع الأقواس المتبعة في الملف.
- تحديد مسار الملف المحول إليه (ملف الترميز XML).

ويتم عرض ملف XML من خلال محررٍ نصوصٍ مثل Notepad++ أو متصفح ويب مثل Firefox أو Chrome أو غيره، كما بالشكل رقم ٤٨:

```xml
<?xml version="1.0" encoding="ISO-8859-1" standalone="no"?>

<corpus xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:noNamespaceSchemaLocation="./schema/TigerXML.xsd"
    id="CRF_SATB" >
    <head external="file:/home/ahmed/TIGERSearchTools/CRF_SATB_generated_header.xml"/>
    <body>
        <s id="s1" >
            <graph root="s1_500" >
                <terminals>
                    <t id="s1_1" word="و" pos="NullPrefix+Conj+NullSuffix_CON" />
                    <t id="s1_2" word="تمل" pos="Present+Active+Verb+NullSuffix_VRB" />
                    <t id="s1_3" word="قيمة" pos="NullPrefix+Noun+Femin+Single_SBJ" />
                    <t id="s1_4" word="كل" pos="NullPrefix+Noun+NullSuffix_IDF" />
                    <t id="s1_5" word="جزء" pos="NullPrefix+Noun+NullSuffix_IDF" />
                    <t id="s1_6" word="100" pos="CD_OBJ" />
                    <t id="s1_7" word="جنيه" pos="NullPrefix+Noun+NullSuffix_TMZ" />
                    <t id="s1_8" word="." pos="PUNC__" />
                </terminals>
                <nonterminals>
                    <nt id="s1_506" cat="NP" >
                        <edge idref="s1_5" label="--" />
                    </nt>
                    <nt id="s1_505" cat="NP" >
                        <edge idref="s1_4" label="--" />
                        <edge idref="s1_506" label="--" />
                    </nt>
                    <nt id="s1_508" cat="NP" >
                        <edge idref="s1_7" label="--" />
                    </nt>
                    <nt id="s1_507" cat="NP" >
                        <edge idref="s1_6" label="--" />
                        <edge idref="s1_508" label="--" />
                    </nt>
                    <nt id="s1_504" cat="NP" >
                        <edge idref="s1_505" label="--" />
                        <edge idref="s1_507" label="--" />
                    </nt>
                    <nt id="s1_503" cat="NP" >
                        <edge idref="s1_3" label="--" />

                        <edge idref="s1_502" label="--" />
                        <edge idref="s1_7" label="--" />
                        <edge idref="s1_505" label="--" />
                        <edge idref="s1_10" label="--" />
                    </nt>
                </nonterminals>
            </graph>
        </s>
    </body>
</corpus>
```

الشكل:٤٨ نموذج من ترميز البنك الشجري بلغة الترميز القابلة للامتداد

وبملاحظة المثال السابق نجد ما يلي:

- تحقق تعريفات نوع الوثيقة DTD – داخل ملف الترميز XML – التي انتهجتها الدراسة.
- تسمية وسم المدونة ID بالبنك الشجري النحوي Syntactic Arabic Treebank (SATB).
- ترقيم العقد النهائية وغير النهائية من الأعلى إلى الأسفل، مبتدئة من الرقم ٥٠٠.
- ترتيب وحدات الجملة ترتيبًا تسلسليًّا.
- يبدأ العنصر بقوسين مثلَّثين < الوسم > وينتهي بقوسين مثلثين، تتوسطهما شرطة مائلة <الوسم />.

•••

الفصل الخامس

توظيف البنك الشجري في بناء النماذج النحوية

وتحقيقًا للغرض من البناء، قامت الدراسة ببناء نموذج نحوي للفئات النحوية الوظيفية، يعتمد على أساليب التعلم الإحصائي في التعرف على الخصائص التركيبية والنحوية بصورة نمطية في المدونة اللغوية الموصفة قيد الدراسة، وذلك من خلال تدريب خوارزمية التصنيف Classification ذات نمط التعلم الموجه Supervised Learning على التوصيفات اللغوية (اليدوية)، التي تعمل على استقراء معطياتها من هذه التوصيفات.

١- منهجية بناء نموذج نحوي للغة العربية الفصحى المعاصرة

قبل البدء في تقديم منهجية بناء النموذج النحوي المنشود، ينبغي أن نقف على مفهوم نمذجة اللغة Language Model في إطار تقنيات الذكاء الاصطناعي ووسائله، فالنمذجة هي الصياغة الصورية للمعرفة اللغوية من خلال استقراء العينة القصدية التي تمثل الواقع اللغوي، ومن ثم كانت النمذجة النحوية هي صورنة درجة الارتباط والتلازم بين المقولات اللغوية المختلفة في المدونة اللغوية الموصَّفة.

وتتمحور الخطوات المنهجية لبناء نموذج نحوي للغة العربية المعاصرة فيما يلي:

١-١ بناء نموذج تركيبي للغة العربية المعاصرة

لما كانت العلاقات التركيبية مدخلًا أساسيًّا لتمثيل الربط والارتباط بين أجزاء الجملة، ودليلًا هاديًا في تحديد الفئات النحوية الوظيفية، فكانت جديرة بأن تكون مصدرًا أساسيا، تستمد منه تقنيات التعلم الآلي الموجه معطياتها؛ للاستدلال على الفئات التصنيفية Classes الموسومة في مدونة البنك الشجري النحوي.

وأفضى ذلك إلى بناء نموذج تركيبي Syntax Model للغة العربية الفصحى المعاصرة لاسيما لغة الصحافة المصرية الإلكترونية، وذلك؛ نظرًا لطبيعة المدونة اللغوية التي يستند إليها مدونة البنك الشجري النحوي، وهذا لا يعني أن هذا النموذج لا يصلح للتطبيق على العربية كلها، وإن كان سيعطي نتيجة أفضل على لغة الصحافة.

وتقوم منهجية بناء نموذج تركيبي إحصائي على ثلاثة إجراءات رئيسية:

١-١-١ المدونة اللغوية

يتطلب بناء النموذج التركيبي في إطار تقنية التعلم الآلي - أول ما يتمثل به - تعيين مدونة للتدريب خالية من الأخطاء الفنية تمامًا، تستمد مصادرها من البنك الشجري النحوي للغة العربية المعاصرة موضوع الدراسة؛ لذا توجب على الدراسة أن تقسم المدونة اللغوية الموصفة عينة الدراسة إلى مدونتين، مدونة للتدريب على الآلية، وأخرى لاختبار هذه الآلية بعد التدريب.

والتزامًا بالشائع في المجتمع المعلوماتي، فقد خصصت الدراسة حوالي 80% من حجم البنك الشجري أي ما يعادل 4,417 جملة للتدريب عن طريق الاختيار العشوائي، و 20% للاختبار أي ما يعادل 1,104 جملة باستخدام خوارزمية K-fold التي يتم بمقتضاها تقيسم المدونة اللغوية بين التدريب والاختبار في عددٍ من الصور المختلفة؛ حتى يمكن تكرار التجربة أكثر من مرة وفقا لعدد صور مدونتي التدريب والاختبار، حيثK تعني عدد صور التقسيم التي تتراوح غالبا ما بين العدد عشرة ١٠ أو العدد خمسة ٥.

ولما كان حجم البنك الشجري النحوي ليس ضخمًا، فقد اقترحت الدراسة العدد K خمسة؛ أي أن تقسّم المدونة اللغوية أو البنك الشجري إلى مدونتين (التدريب والاختبار) في خمس صور مختلفة.

١-١-٢ خوارزمية التدريب

لما كان تطبيق خوارزميات التحليل التركيبي يتطلب الاعتماد على منصة أو بيئة حاسوبية، فقد اقترحت الدراسة الاعتماد على منصة ستانفورد البرمجية؛ لما تتوافر فيها العديد من الخيارات التي تسمح بتفعيل أكثر من خوارزمية في وقت واحد، ومن هذه الخوارزميات:

- **خوارزمية** Probabilistic Context Free Grammar (PCFG)

هي إحدى الخوارزميات المستخدمة في التحليل التركيبي الآلي، تقوم فكرتها على تحليل الجملة المدخلة عن طريق بناء الشجرة التحليلية من الأعلى أي جذر الشجرة (وهو الرمز S) للأسفل أي الكلمات، وذلك في إطار تطبيق القوانين الاحتمالية الشرطية على قوانين بنية العبارة التي تم استخلاصها من مدونة التدريب.[١]

ثمة مشكلات قد تطرأ على تطبيق هذه الخوارزمية خاصة على اللغة العربية، منها:

١- **مشكلة التكرار** من ناحية اليسار التي قد تحدث أثناء تطبيق قوانين العقد اللانهائية في أعلى الشجرة؛ مما تؤدي إلى حلقات تكرارية غير نهائية Infinite loops، ومن ثم فشلها في تحليل أجزاء في القوانين. مثال ذلك قانون: VP→VP PP PP حيث نجد تكرارًا للقانون من ناحية اليسار، ووفقًا لطريقة البحث التي تتبعها الخوارزمية بتحليل الجزء الأيسر أولًا، نجد أنها تدخل في حلقة تكرارية غير نهائية من توليد الأشجار.

٢- **مشكلة الغموض** أي عدم قدرة الخوارزمية على التعامل مع الجمل الغامضة التي تحتمل تحليلين أو أكثر. مثال ذلك جملة "رأيت ولدًا بنظارات" التي تحتمل معنيين يحددهما السياق النحوي.

- **خوارزمية** Cocke–Younger–Kasami CYK

وتفاديًا للمشكلات التي ظهرت في التحليل من الأعلى إلى الأسفل Top-down Parsing، قدم كل من الأمريكي جون كوك John Cocke، ودانيال ينجر Daniel Younger، والياباني تاداو كاسامي Tadao Kasami فكرة جديدة تنطلق في التحليل التركيبي من الأسفل أي الكلمات إلى الأعلى Bottom-up parsing أي جذر الشجرة مع مراعاة الاعتماديات بين كلمات الشجرة في الجملة.[٢]

---

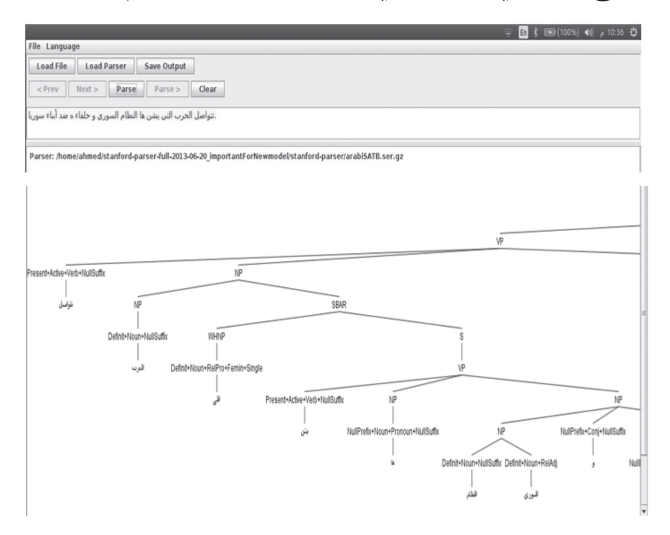
(1) Jurafsky.D & Martin.J. (2007). Ibid.P.478.

(2) Jurafsky.D & Martin.J. (2007). Ibid.P.486.

وتتجه هذه الخوارزمية - أولا - إلى تعيين قوانين العقد النهائية لمدونة التدريب، ثم تتوسع في إيجاد كل العقد غير النهائية الممكنة لتغطية المستوى الثاني، ثم حساب احتمالات أفضل عقدة مناسبة للعقد النهائية، ثم تتوالى هذه العملية في كل المستويات إلى أن تصل إلى جذر الشجرة S.

وقد اعتمدت الدراسة على خوارزمية CYK - باعتبارها أفضل خوارزميات التحليل التركيبي المتاحة - في ضوء منصة ستانفورد التي تستخلص القوانين من مدونة التدريب في صيغة تشومسكي العادية Chomsky normal form (CNF) التي تنظم كل القوانين في جانبين[1]، بحيث يكون لديها قابلية التعامل مع الخوارزمية.

وبعد التدريب وإقامة النموذج التركيبي الإحصائي، يمكننا الآن تجريبه أي النموذج - من حيث التفعيل - من خلال إدخال جملة " تتواصل الحرب التي يشنها النظام السوري وحلفاء هـ ضد أبناء سوريا"[2] عبر واجهة الرسوم GUI، ومن ثم تحليلها باستخدام النموذج التركيبي الإحصائي، كما يظهر بالشكل رقم ٤٩.



الشكل:٤٩ تحليل جملة" تتواصل الحرب..." عن طريق النموذج التركيبي الإحصائي

ولعدم وضوح الرؤية كاملة للهيئة الشجرية للجملة المحللة في المنصة الحاسوبية،

---

(١) الجانب الأيسر هو المركب، والجانب الأيمن هو المكون من مركبين أو كلمة.

(٢) بالطبع ستتم التجزئة آليًا حال إدخال المستخدم للنص المراد تحليله تركيبيًّا، أي بالإمكان إدماج أداة التجزئة بمنصة التحليل التركيبي.

يمكننا أن نعرضها من خلال التمثيل البديل (الأقواس).

"تتواصل الحرب التي يشنها النظام السوري وحلفائه ضد أبناء سوريا." إسلام أون لاين ٢٠١٦.

```
(ROOT
  (S
    (VP (Present+Active+Verb+NullSuffix تـتـواصل)
      (NP
        (NP (Definit+Noun+NullSuffix الـحرب))
        (SBAR
          (WHNP (Definit+Noun+RelPro+Femin+Single الـتي))
          (S
            (VP (Present+Active+Verb+NullSuffix يـشن)
              (NP (NullPrefix+Noun+Pronoun+NullSuffix ها))
              (NP
                (NP (Definit+Noun+NullSuffix الـنظام) (Definit+Noun+RelAdj
الـسوري))
                (NullPrefix+Conj+NullSuffix و)
                (NP (NullPrefix+Noun+NoSARF+Plural+NullSuffix حلـفاء)
                  (NP (NullPrefix+Noun+Pronoun+NullSuffix ـه))))))))
      (NP (NullPrefix+Noun+NullSuffix ضد)
        (NP (NullPrefix+Noun+Plural+NullSuffix أبـناء)
          (NP (NullPrefix+Noun+NoSARF+Femin+Single سـوريا)))))
    (PUNC .)))
```

١-١-٣ التقييم

نستطيع أن نقيم دقة الآلية أو النموذج الإحصائي من خلال مدونة الاختبار باعتبارها معيارًا ذهبيًّا Gold Standard مشابهًا لمدونة التدريب، وليست مطابقة لها. وقد استعانت الدراسة بأداة تقييم التحليل المكوني evalC[١] للباحث الأسكتلندي فدريكو سنجاتي Federico Sangati بجامعة أدنبرة Edinburgh، فكانت النتائج كما يلي:

(١) يمكن تحميل الأداة من خلال هذا الرابط:
evalC. (2010). http://homepages.inf.ed.ac.uk/fsangati/

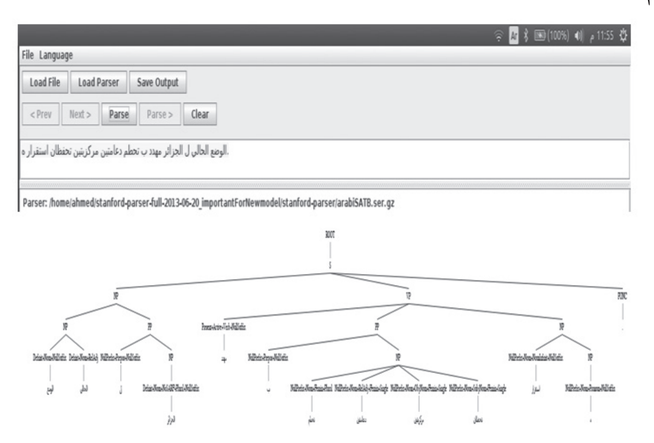
طول الجملة	تقييم التحليل المكوناتي Constituency parsing		
أقل أو يساوي 70 كلمة.	الدقة Precision	المراجعة Recall	المتوسط التوافقي F-measure
	%81.37	81.21%	%81.29

الجدول:٣٦ تقييم النموذج التركيبي الإحصائي على مستوى الوحدات

وقطعا تختلف النتائج للأفضل عندما يكون المدخل معنونًا بالأقسام الكلامية باستخدام المعنون الآلي، إذ كلما كانت دقة توصيفات الأقسام الكلامية عالية، كانت عونًا للخوارزمية على دقة تحديد العقد Nodes المناسبة.

ولإثبات الافتراض الذي افترضه البحث حول تحسن النتائج بعنونة النص المدخل مسبقًا باستخدام آلية عنونة الأقسام الكلامية، فقد تم اختيار جملة عشوائية من الصحافة المصرية، وقد أوجدناها في صورتين، تحتوي الأولى منها على النص الخام، وتحتوي الأخرى على النص نفسه معنونًا بالأقسام الكلامية.

كانت نتائج الصورة الأولى بعد تمريرها على النموذج التركيبي الإحصائي كما بالشكل رقم ٥٠:



الشكل: ٥٠ تحليل جملة " الوضع الحالي للجزائر..." عن طريق النموذج التركيبي الإحصائي

ولعدم وضوح الرؤية للهيئة الشجرية للجملة المحللة في المنصة الحاسوبية، يمكننا أن نعرضها من خلال التمثيل البديل (الأقواس).

"الوضع الحالي للجزائر مهدد بتحطم دعامتين مركزيتين تحفظان استقراره."

```
(ROOT
  (S
    (NP
      (NP (Definit+Noun+NullSuffix الوضع) (Definit+Noun+RelAdj الحالي))
      (PP (NullPrefix+Prepos+NullSuffix ل)
        (NP (Definit+Noun+NoSARF+Plural+NullSuffix الجزائر))))
    (VP (Present+Active+Verb+NullSuffix مهدد)
      (PP (NullPrefix+Prepos+NullSuffix ب)
        (NP (NullPrefix+Noun+Femin+Plural تحطم) (NullPrefix+Noun+RelAdj
+Femin+Single دعامتين) (NullPrefix+Noun+ObjNoun+Femin+Single مركزيتين)
(NullPrefix+Noun+SubjNoun+Femin+Single تحفظان)))
      (NP (NullPrefix+Noun+NounInfinit+NullSuffix استقرار)
        (NP (NullPrefix+Noun+Pronoun+NullSuffix ـه))))

    (PUNC .)))
```

وتتطلب الصورة الثانية للنص التي تتضمن العنونة بالأقسام الكلامية تنسيقًا معينًا، كما يظهر في الشكل التالي.

```
الوضع/Definit+Noun+NullSuffix الحالي/Definit+Noun+RelAdj ل/
NullPrefix+Prepos+NullSuffix الجزائر/Definit+Noun+NoSARF+Pl
ural+NullSuffix مهدد/NullPrefix+Noun+SubjNoun+NullSuffix ب/
NullPrefix+Prepos+NullSuffix تحطم/NullPrefix+Noun+NounInfini
t+NullSuffix دعامتين/NullPrefix+Noun+Femin+Binary+NonAdjun
ct+MANSS_MAGR مركزيتين/NullPrefix+Noun+Femin+Binary+NonAd
junct+MANSS_MAGR تحفظان/Present+Passive+Verb+MARF+SubjP
ro استقرار/NullPrefix+Noun+NounInfinit+NullSuffix ـه/NullPr
efix+Noun+Pronoun+NullSuffix ./PUNC
```

وكانت نتائج الصورة الثانية بعد تمريرها على النموذج التركيبي الإحصائي كما بالشكل التالي:

"الوضع الحالي للجزائر مهدد بتحطم دعامتين مركزيتين تحفظان استقراره."

```
(ROOT
  (S
    (S
      (NP
        (NP (Definit+Noun+NullSuffix الوضع)
(Definit+Noun+RelAdj الحالي))
        (PP (NullPrefix+Prepos+NullSuffix ل)
          (NP (Definit+Noun+NoSARF+Plural+NullSuffix
الجزائر))))
      (NP (NullPrefix+Noun+SubjNoun+NullSuffix مهدد)
        (PP (NullPrefix+Prepos+NullSuffix ب)
          (NP (NullPrefix+Noun+NounInfinit+NullSuffix
تحطم))))
            (NullPrefix+Noun+Femin+Binary+NonAdjunct+MA
```

```
NSS دعامتين) (NullPrefix+Noun+Femin+Binary+NonAdjunct+MAN
SS مركزيتين)
     (S
       (VP (Present+Passive+Verb+MARF+SubjPro تحفظان)
         (NP (NullPrefix+Noun+NounInfinit+NullSuffix
استقرار))
         (NP (NullPrefix+Noun+Pronoun+NullSuffix ه))))
     (PUNC .)))
```

والناظر في الصورتين بعد تمريرهما على النموذج التركيبي الإحصائي، سيجد اختلافًا واضحا في التحليل، ففي الصورة الأولى نجد كلمة "مهدد" رأسا للمركب الفعلي، وكذلك كلمة " تحفظان" مركبا اسميا، وهذا ناتج عن عدم دقة العنونة بالأقسام الكلامية التي يعينها النموذج التركيبي، أما في الصورة الثانية، فنجد أن كلمة "مهدد" رأسا للمركب الاسمي، وكلمة " تحفظان" رأسا لمركب فعلي داخل جملة صغرى، وهذا ناتج عن دقة العنونة بالأقسام الكلامية المسبقة.

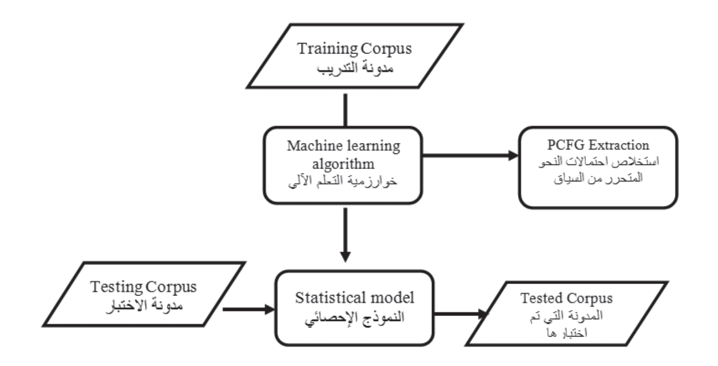
تعد دقة العنونة بالأقسام الكلامية دليلًا هاديًا في تحديد الكلمات داخل مكوناتها، ومن ثم انتظام المكونات / المركبات داخل الجملة.

ويمكننا - أيضا- تقييم النموذج التركيبي الإحصائي بصورة أخرى من خلال تمرير نصوص مدونة الاختبار على معنوِّن الأقسام الكلامية (السراج)، ومن ثم إدخال النصوص معنونة بالأقسام الكلامية إلى النموذج التركيبي الإحصائي. وباستعانة أداة تقييم التحليل المكوني evalC كانت نتائج التقييم كالآتي:

طول الجملة	تقييم التحليل المكوناتي Constituency parsing		
	الدقة Precision	المراجعة Recall	المتوسط التوافقي F-measure
أقل أو يساوي 70 كلمة.	83.83%	82.20%	%83.00

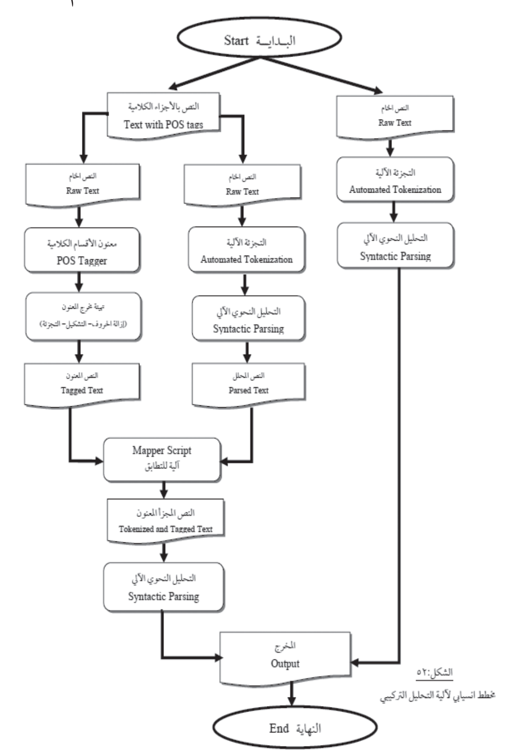
الجدول:٣٧ تقييم النموذج التركيبي الإحصائي على مستوى الوحدات حال إدخال النص معنونًا - آليًّا - بالأقسام الكلامية

ويوضح الشكل رقم ٥١ التخطيط الانسيابي لبناء النموذج التركيبي الإحصائي.



الشكل: ٥١ مخطط انسيابي لبناء النموذج التركيبي

ويلخص التخطيط الانسيابي التالي الخطوات الإجرائية لتنفيذ آلية التحليل التركيبي سواء أكان المدخل نصًّا فقط أو نصًّا معنونًا – آليًّا – بالأقسام الكلامية.



الشكل: ٥٢ مخطط انسيابي لآلية التحليل التركيبي

١-٢ بناء نموذج نحوي تصنيفي للغة العربية المعاصرة

وسعيًا لتوفير المعطيات اللازمة للتحليل اللغوي الأعمق للفهم الأتوماتي للنصوص اللغوية، فكان الانطلاق من البيانات الوظيفية لوحدات الجمل - بتحليلها وإدراك علاقتها - خطوة أساسية في استشفاف المعلومات التي بدورها تعين على عملية الفهم الآلي للغات الطبيعية.

وتقوم منهجية بناء نموذج نحوي (للفئات النحوية الوظيفية) إحصائي على ثلاثة إجراءات رئيسية:

١-٢-١ المدونة اللغوية (استخلاص السمات من البنك الشجري النحوي)

ولما كان الإدراك الآلي للفئات أو التصنيفات Labels في إطار تقنيات الذكاء الاصطناعي يتطلب أدلة هادية، ممثلة في عدد محدودٍ من الخصائص المميزة التي تعين المصنف على الاستنتاج، فقد اقترحت الدراسة سمات الفئة من حيث موقعها في البناء الشجري[١]، بالإضافة إلى الأقسام الكلامية التي تنتمي إليها.

وقد تمثلت السمات Features التي اقترحتها الدراسة لاستخلاصها من البناء الشجري على الترتيب التالي:

- الأقسام الكلامية.
- الفئة التركيبية للكلمة من حيث كونها مركبا اسميا أو فعليا،....
- وضْع الكلمة في البناء الشجري من حيث التسلسل العائلي (الأب - الأخوات).
- وضْع الكلمة في هيكلة السين البارية من حيث الرأس والمكمل.

ولما كانت خوارزميات التصنيف الآلي تتعامل مع المعطيات أو السمات بصورة رياضية، بحيث يمكن حساب علاقة مصفوفات ومتجهات الوثيقة Matrix and Vector من خلال بناء نموذج جدولة الكلمات Bag of Words للوثيقة، فكان النظام

(١) هناك منهج شائع بين العديد من اللغات لاستخلاص السمات من البنية الشجرية، يطلق عليه IOB، إلا أنه يقتصر فقط على المركبات؛ إذ يحدد نوع كلمة من حيث وقوعها داخل المركب I-insight token or phrase أو خروجها عن هذا المركب O-outsight token or phrase أو وقوعها في بداية المركب bigen token or phrase-B ينظر:
Attardi, G. & Dell'Orletta, F. (2008). Chunking and Dependency Parsing. In Proc of LREC Workshop on Partial Parsing: Between Chunking and Deep Parsing. Marrakech, Morocco

الجدولي لعرض معطيات التدريب هو الأنسب من حيث ملاءمته لطبيعة الخوارزميات الرياضية. ويوضح الجدول رقم ٣٨ وضع معطيات التدريب التي اقترحتها الدراسة.

Class[1] الفئة النحوية الوظيفية	X-Bar Representations وضع الكلمة في تمثيل السين البارية	Parent brother أخت الكلمة في التسلسل الشجري	Parent father أب الكلمة في التسلسل الشجري	Phrase Cat فئة المركب	POS الأقسام الكلامية	Word الكلمة

الجدول:٣٨ وضع السمات في بنية جدولية

ولما اقتضت الحاجة إلى استخلاص هذه المعلومات آليا من التحليل الشجري، فكان جدير بأن يتم الاستخلاص من البناء الشجري الموصف بلغة الترميز القابلة للامتداد XML، إذ تتوافق هذه السمات من الناحية الترتيبية – إلى حد ما – مع التنظيم الهيكلي لملف XML، ومن ثم تكون عملية استخلاص السمات أكثر دقة وإحكامًا.

وقد تم تطوير أداة استخلاص السمات Feature Extraction من البناء الشجري – المرمز بلغة الترميز القابلة للامتداد – باستخدام لغة الجافا البرمجية، مدعمة العديد من أنظمة التشغيل.

ولتوضيح عمل آلية استخلاص السمات من البناء، يمكننا عرض مدخلات الآلية ومخرجاتها، فيعرض الشكل التالي مدخلات الآلية المتمثلة في البنية الشجرية المرمزة بلغة الترميز القابلة للامتداد.

(١) راجع الفئات الوظيفية ص ٢١٢.

```xml
<?xml version="1.0" encoding="ISO-8859-1" standalone="no"?>

<corpus xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:noNamespaceSchemaLocation="./schema/TigerXML.xsd"
    id="CRF_SATB" >
    <head external="file:/home/ahmed/TIGERSearchTools/CRF_SATB_generated_header.xml"/>
    <body>
        <s id="s1" >
            <graph root="s1_500" >
                <terminals>
                    <t id="s1_1" word="و" pos="NullPrefix+Conj+NullSuffix_CON" />
                    <t id="s1_2" word="تمل" pos="Present+Active+Verb+NullSuffix_VRB" />
                    <t id="s1_3" word="قيمة" pos="NullPrefix+Noun+Femin+Single_SBJ" />
                    <t id="s1_4" word="كل" pos="NullPrefix+Noun+NullSuffix_IDF" />
                    <t id="s1_5" word="جزء" pos="NullPrefix+Noun+NullSuffix_IDF" />
                    <t id="s1_6" word="100" pos="CD_OBJ" />
                    <t id="s1_7" word="جنيه" pos="NullPrefix+Noun+NullSuffix_TMZ" />
                    <t id="s1_8" word="." pos="PUNC_" />
                </terminals>
                <nonterminals>
                    <nt id="s1_506" cat="NP" >
                        <edge idref="s1_5" label="--" />
                    </nt>
                    <nt id="s1_505" cat="NP" >
                        <edge idref="s1_4" label="--" />
                        <edge idref="s1_506" label="--" />
                    </nt>
                    <nt id="s1_508" cat="NP" >
                        <edge idref="s1_7" label="--" />
                    </nt>
                    <nt id="s1_507" cat="NP" >
                        <edge idref="s1_6" label="--" />
                        <edge idref="s1_508" label="--" />
                    </nt>
                    <nt id="s1_504" cat="NP" >
                        <edge idref="s1_505" label="--" />
                        <edge idref="s1_507" label="--" />
                    </nt>
                    <nt id="s1_503" cat="NP" >
                        <edge idref="s1_3" label="--" />

                        <edge idref="s1_504" label="--" />
                    </nt>
                    <nt id="s1_502" cat="VP" >
                        <edge idref="s1_2" label="--" />
                        <edge idref="s1_503" label="--" />
                    </nt>
                    <nt id="s1_501" cat="S" >
                        <edge idref="s1_1" label="--" />
                        <edge idref="s1_502" label="--" />
                        <edge idref="s1_8" label="--" />
                    </nt>
                    <nt id="s1_500" cat="ROOT" >
                        <edge idref="s1_501" label="--" />
                    </nt>
                </nonterminals>
            </graph>
        </s>
```


الشكل:٥٣ نموذج لمدخل آلية استخلاص السمات

و	NullPrefix+Conj+NullSuffix	S	S_501	VP_502+.	Head	CON
تمل	Present+Active+Verb+NullSuffix	VP	VP_502	NP_503	Head	VRB
قيمة	NullPrefix+Noun+Femin+Single	NP	NP_503	NP_504	Head	SBJ
كل	NullPrefix+Noun+NullSuffix	NP	NP_505	NP_506	Head	IDF
جزء	NullPrefix+Noun+NullSuffix	NP	NP_506	_	_	IDF
100	CD	NP	NP_507	NP_508	Head	OBJ
جنيه	NullPrefix+Noun+NullSuffix	NP	NP_508	_	_	TMZ
.	PUNC	S	S_501	w+VP_502	Head	MRK

الجدول:٣٩ نموذج لمخرجات آلية استخلاص السمات

وبذلك تتوفر مدونة التدريب Training set التي تخضع لوسائل التعلم الآلي الموجه لاسيما المصنفات الآلية Classifiers.

١-٢-٢ خوارزمية التدريب (التصنيف)

ثمة العديد من المصنفات المستخدمة في بناء التصنيفات الآلية، ومنها على سبيل المثال لا الحصر:

- آلات الدعم الموجهة SVMs
- مصنف الحقول العشوائية المشروطة CRF
- مصنف أشجار القرار DT

والفارق بين هذه المصنفات هو النظام الرياضي المستخدم - داخل التقنية - لحل إشكالية التصنيف Classification. وفي ضوء الهدف المنشود، اقترحت الدراسة استخدام مصنف الحقول العشوائية المشروط[١] ++CRF [٢] الذي يعتمد على فكرة التسلسل Sequence الإحصائي الاحتمالي التي يتم بمقتضاها حساب أثر التلازم بين الكلمات أو السمات في الجملة، ومن ثم استخلاص المقاييس Weights الحاكمة التي يتم من خلالها إدراك الفئات Classes.[٣]

كما يمتاز هذا المصنف بخاصية إضافية عن المصنفات الأخرى، وهي القالب Template الذي يمكن من خلاله تعيين سمات إضافية في مصفوفة السمات عن طريق استخدام (الماكرو) المتمثل في هذا النمط x [row,col] % أي تعيين سمات السمات من خلال فكرة التجاور بين الكلمات أو سماتها في الصفوف والأعمدة.

---

(١) تتشابه هذه الحقول العشوائية المشروطة مع نموذج ماركوف الخفي HMM، إلا أن الأخير يعتمد في تنبؤه على الحالات غير الملحوظة unobserved (hidden) states، أما الحقول العشوائية المشروطة فتعتمد على البيانات الملحوظة observed Data أي المصفوفة الممثلة في هيئة أعمدة وصفوف. انظر:
Ponomareva N, Rosso P, Pla F, Molina A. (2007). Conditional random fields vs. hidden markov models in a biomedical named entity recognition task. In Proc. of Int. Conf. Recent Advances in Natural Language Processing, RANLP.

(٢) يمكن تحميل الأداة من خلال هذا الرابط:
CRF++. (2003). https://taku910.github.io/crfpp/

(3) Sutton, Ch. & McCallum, A. (2010). An Introduction to Conditional Random Fields. Cornell University Library.P.2.

وتتضح الفكرة من خلال عرض القالب المدرج في مصنف الحقول العشوائية المشروطة الذي اعتمدته الدراسة في تعيين سمات الكلمات المرتبطة بسماتها.[(١)]

```
# Unigram
U00:%x[-2,0]
U01:%x[-1,0]
U02:%x[0,0]
U03:%x[1,0]
U04:%x[2,0]
U05:%x[-1,0]/%x[0,0]
U06:%x[0,0]/%x[1,0]

U10:%x[-2,1]
U11:%x[-1,1]
U12:%x[0,1]
U13:%x[1,1]
U14:%x[2,1]
U15:%x[-2,1]/%x[-1,1]
U16:%x[-1,1]/%x[0,1]
U17:%x[0,1]/%x[1,1]
U18:%x[1,1]/%x[2,1]
U20:%x[-2,1]/%x[-1,1]/%x[0,1]
U21:%x[-1,1]/%x[0,1]/%x[1,1]
U22:%x[0,1]/%x[1,1]/%x[2,1]

# Bigram
B
```

(١) مع العلم يمكن تغيير وضع السمات في القالب على المستويين (الصف، العمود).

يلاحظ أن القالب يتدرج في تعيين السمات، حيث نجده في المجموعة الأولى يتمحور حول سمات الكلمة من حيث ما يسبقها أو يتلوها؛ إذ يشغل الصف - وفقًا للنمط - الأعداد الموجبة التي تنم عن علاقة الكلمة بما بعدها، و الأعدد السالبة التي تنم عن علاقة الكلمة بما قبلها، أما العمود فيلزم الحالة الصفرية هنا، ثم تتدرج هذه الحالة بالتوسع في المجموعة التالية؛ لتشمل الصف والعمود معًا، وهكذا...

وبعد التدريب وإقامة نموذج المصنف الآلي، يمكننا تجريبه من خلال إدخال جملة " تتواصل الحرب التي يشنها النظام السوري وحلفاء هـ ضد أبناء سوريا" المحللة تركيبيًّا عن طريق النموذج التركيبي الإحصائي:

```
(ROOT
  (S
    (VP (Present+Active+Verb+NullSuffix تتواصل)
      (NP
        (NP (Definit+Noun+NullSuffix الحرب))
        (SBAR
          (WHNP (Definit+Noun+RelPro+Femin+Single التي))
          (S
            (VP (Present+Active+Verb+NullSuffix يشن)
              (NP (NullPrefix+Noun+Pronoun+NullSuffix ها))
              (NP
                (NP (Definit+Noun+NullSuffix النظام) (Definit+Noun+RelAdj
السوري))
                (NullPrefix+Conj+NullSuffix و)
                (NP (NullPrefix+Noun+NoSARF+Plural+NullSuffix حلفاء)
                  (NP (NullPrefix+Noun+Pronoun+NullSuffix ه))))))))
      (NP (NullPrefix+Noun+NullSuffix ضد)
        (NP (NullPrefix+Noun+Plural+NullSuffix أبناء)
          (NP (NullPrefix+Noun+NoSARF+Femin+Single سوريا)))))
    (PUNC .)))
```

فكانت النتائج كما يلي، مشتملة على الوحدات (بنظام النقل الكتابي باكولتر)[١] والسمات والفئات الوظيفية.

```
1  ttwASl   Present+Active+Verb+NullSuffix   VP   VP_502   NP_503+NP_514   Head   VRB
2  AlHrb   Definit+Noun+NullSuffix   NP   NP_504   _   _   SBJ
3  Alty   Definit+Noun+RelPro+Femin+Single   WHNP   WHNP_506   _   _   CON
4  y$n   Present+Active+Verb+NullSuffix   VP   VP_508   NP_509+NP_510   Head   VRB
5  hA   NullPrefix+Noun+Pronoun+NullSuffix   NP   NP_509   _   _   OBJ
6  AlnZAm   Definit+Noun+NullSuffix   NP   NP_511   Alswry   _   SBJ
7  Alswry   Definit+Noun+RelAdj   NP   NP_511   AlnZAm   _   MOD
8  w   NullPrefix+Conj+NullSuffix   NP   NP_510   NP_511+NP_512   Head   CON
9  HlfA'   NullPrefix+Noun+NoSARF+Plural+NullSuffix   NP   NP_512   NP_513   Head   NN
10 h   NullPrefix+Noun+Pronoun+NullSuffix   NP   NP_513   _   _   IDF
11 Dd   NullPrefix+Noun+NullSuffix   NP   NP_514   NP_515   Head   MOD
12 >bnA'   NullPrefix+Noun+Plural+NullSuffix   NP   NP_515   NP_516   Head   IDF
13 swryA   NullPrefix+Noun+NoSARF+Femin+Single   NP   NP_516   _   _   IDF
14 .   PUNC   S   S_501   VP_502   Head   MRK
15
```

الجدول: ٤٠ نموذج لمخرجات آلية التصنيف الآلي لجملة" تتواصل الحرب..."

وبتحديد الوحدات والفئات الوظيفية دون السمات:

```
1  ttwASl   VRB
2  AlHrb   SBJ
3  Alty   CON
4  y$n   VRB
5  hA   OBJ
6  AlnZAm   SBJ
7  Alswry   MOD
8  w   CON
9  HlfA'   NN
10 h   IDF
11 Dd   MOD
12 >bnA'   IDF
13 swryA   IDF
14 .   MRK
15
```

الجدول: ٤١ تحديد الفئات النحوية الوظيفية دون السمات في مخرجات آلية التصنيف الآلي لجملة" تتواصل الحرب..."

(١) قد واجهت مشكلة في استخدام الحروف العربية في أداة ++CRF؛ لذا أدمجت في الأداة محول النقل الكتابي باكولتر لتحويل النص المدخل إلى نظام باكولتر، ومن ثم يكون المخرج بنظام هذا النقل الكتابي، ومع ذلك يمكن تحويل هذا النظام ثانيا إلى الحروف العربية.

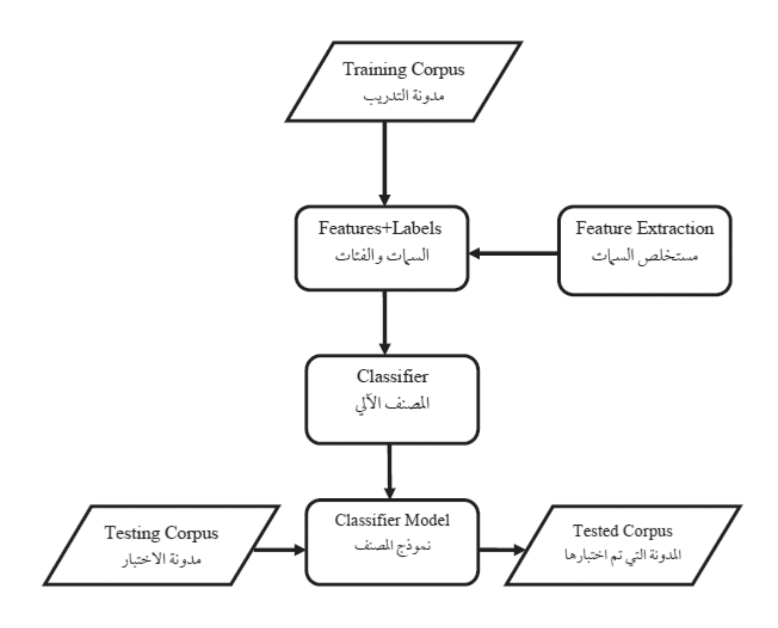
١-٢-٣ التقييم

وبعد تدريب مصنف الحقول العشوائية على المعطيات اللغوية التي تم استخلاصها من البنية الشجرية، وإقامة النموذج النحوي الإحصائي، يمكننا أن نقيّم هذا النموذج من خلال مدونة الاختبار، بحساب معدل الخطأ للفئة Class error Rate (CER) الذي يتم في إطار المقارنة بين التحليل الآلي ومدونة الاختبار (المعيار الذهبي).

عدد الكلمات	عدد الوحدات	عدد الجمل	حجم التدريب	حجم الاختبار	معدل الخطأ للفئة
105,781	117,785	5,521	4,417	1,104	10,4%

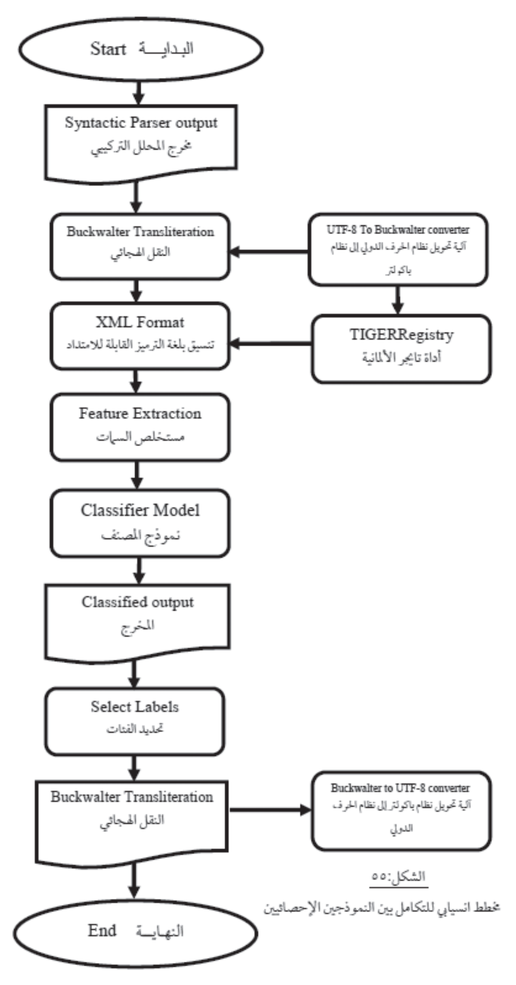
الجدول:٤٢ إحصائيات النموذج التصنيفي وتقييمه

ويوضح الشكل رقم ٥٤ التخطيط الانسيابي لبناء نموذج المصنف الآلي للفئات الوظيفية:



الشكل:٥٤ مخطط انسيابي لبناء النموذج التصنيفي

وبعد الانتهاء من بناء النموذجين الإحصائيين بصورة منفردة لكل منهما، يمكننا استعراض التكامل فيما بينهما – ليتحقق الهدف المنشود – من خلال التخطيط الانسيابي التالي:



الشكل:٥٥ مخطط انسيابي للتكامل بين النموذجين الإحصائيين

٣-١ التقييم الشامل لتكامل النموذجين الإحصائيين (النموذج النحوي)

وبتقييم هذه النموذج النحوي المتكامل أي عندما يستند إلى معطيات التحليل التركيبي الآلي لاستخلاص السمات التي تعينه على التعرف على الفئات الوظيفية كما في المخطط السابق، يمكننا الاعتماد على نصوص مدونة الاختبار المعنونة بالأقسام

الكلامية آليًّا - التي تم اعتمادها سابقا في تقييم النموذج التركيبي الإحصائي - في تقييم النموذج باعتبارها مدخلًا للنموذج التركيبي الإحصائي ثم آلية استخلاص السمات ثم النموذج التصنيفي.

وبعد تمرير نصوص مدونة الاختبار على الخطوات الإجرائية لمنهجية بناء النموذج النحوي للغة العربية المعاصرة، ثم مقارنة نتائجها بمدونة الاختبار (المعيار الذهبي)، كانت النتائج كما يلي:

الاختبار	عدد الجمل	دقة النموذج التركيبي	معدل الخطأ للفئة
إحصائيات	1,104	83,00	12,3

الجدول:٤٣ تقييم النموذج النحوي المتكامل

•••

الخاتمة والنتائج

الخاتمة

حاول هذا الكتاب أن يقدم الإطار المنهجي لبناء البنك الشجري النحوي من خلال التطبيق على نصوص من اللغة العربية المعاصرة، ثم موظفًا إياه في بناء نموذج نحوي إحصائي باستخدام تقنيات التعلم الآلي (خوارزمية التصنيف). وقد أمكن من خلال البناء والتوظيف التوصل إلى عدد من النتائج، كان من أهمها ما يلي:

بلغ عدد كلمات المدونة اللغوية 105,781 كلمة - من نصوص موقع إسلام أون لاين - بعد إزالة الفواصل والمساحات الفارغة، والكلمات المكتوبة بحروف إنجليزية، تضمها 110 وثيقة نصيّة.

بلغ عدد الكلمات غير المكررة في المدونة اللغوية 38,321 كلمة. وبالنظر إلى إحصاء (الكلمات - الكلمات غير المتكررة) في المدونة اللغوية، نجد أن هناك علاقة شبه ثابتة بين حجم النص، وعدد المفردات غير المتكررة فيه. وهذا ما يؤكد تنوع الموضوعات في العينة المختارة (المدونة اللغوية) من المجتمع اللغوي غير المحدود.

تنوعت نصوص المدونة اللغوية في عدد من المجالات (السياسة - الاقتصاد - الفنون - الاجتماع - العلوم - الرياضة) بهدف تحقيق قدرٍ كافٍ من أنماط التركيب الجملي في الاستعمال، والمفردات والأساليب المختلفة.

أمكن تهيئة المدونة اللغوية للمعالجة الآلية، بحيث يكون لديها قابلية للتعامل مع الأساليب البرمجية للمعالجة الآلية للغة الطبيعية، وقد أفاد منها الباحث في التحليل الآلي عبر تقنيات اللغة، وكذلك في بناء النماذج الإحصائية.

أمكن إخضاع النصوص لعملية التجزئة؛ لتمكينها من التحليل الآلي، والعنونة اللغوية، فبلغ عدد جمل المدونة اللغوية بعد التجزئة 5521 جملة، وعدد الوحدات Tokens 117,785 وحدة.

أمكن عنونة نصوص المدونة اللغوية بالأقسام الكلامية؛ لتكونَ مدخلًا أساسيًّا لتعيين العلاقات التركيبية فيما بعد.

أمكن عنونة نصوص المدونة اللغوية بالعلاقات التركيبية والفئات الوظيفية؛ ليكتمل البناء الشجري، ويكون موردًا لغويًّا لتطبيقات معالجة اللغة العربية، وأداة للبحث في الظواهر اللغوية.

إن أية عملية تحليل نحوي تتوخى الطرق المنهجية في الوصف النحوي، لابد أن تتوفر لديها ثلاثة قيود أساسية (التمثيل النحوي – المحتوى النحوي – النظرية النحوية).

أمكن اختبار دقة البناء الشجري – من خلال القياس الإحصائي لتعيين درجة المصداقية أو التوافق بين تحليلات العينة المختلفة inter-rater agreement - باختيار عينة عشوائية ١٠٠ جملة قد تم تحليها من قبل، ثم إعادة تحليلها مرة أخرى من قبل لغوي آخر قد تدرب على منهجية التحليل، بحيث يتحقق ركنا المعادلة لقياس نسبة التوافق. وقد أمكن مقارنتها باستخدام أداة evalC، فبلغت نسبة التوافق بين التحليلين ٩٥٪. وكان جل الاختلاف الوارد في العينتين في مواضع تعيين عناصر البناء الشجري للبدل والصفة والإضافة؛ لتقارب البناء فيهما بينهم.

أمكن البحث في الظواهر اللغوية لا سيما أنماط التركيب الجملي – باستخدام أداة Tregex[(١)] – التي تم توصيفها في إطار مدونة لغوية تعكس الواقع اللغوي، مع معرفة ترددات تلك الظواهر أو الأنماط؛ وذلك للإفادة منها في تعليم اللغة العربية للناطقين بها أو بغيرها.

(١) يمكن تحميل هذه الأداة من خلال هذا الرابط:
Tregex. (2005). http://nlp.stanford.edu/software/tregex.shtml

أمكن ترميز البنك الشجري بلغة الترميز القابلة للامتداد XML؛ لسهولة استدعاء المعلومات اللغوية أو دمجها في التقنيات البرمجية.

بناء نموذج تركيبي للغة العربية المعاصرة، باستخدام تقنيات التعلم الآلي (خوارزمية CYK)، وقد بلغت دقة نتائجه 81,29% حال إذا كان المدخل نصًّا خامًا Raw Text، أما إذا كان المدخل نصًّا معنونًا - آليًّا - بالأقسام الكلامية، فقد بلغت دقة نتائجه 83,00%.

تبيّن البحث كلما ازدادت دقة توصيفات الأقسام الكلامية للنص المراد تحليله تركيبيًّا، ازدادت دقة خوارزمية التحليل التركيبي على تحديد العقد Nodes المناسبة لأجزاء الجملة.

بناء نموذج نحوي وظيفي للغة العربية المعاصرة - باستخدام آلية التصنيف (CRF++Classifier) - يقوم على المفهوم النحوي Syntactic Concept في تصنيف الفئات النحوية؛ وذلك للإفادة منه في تطبيقات السؤال والجواب، وغيرها من التطبيقات التي تقوم علي الفهم الأتوماتي للنصوص اللغوية. وقد بلغ معدل الخطأ في نتائجه أثناء اختباره 10,4%.

•••

المراجع

المراجع

أوَّلًا: المراجع العربية:

- ابن جني (أبو الفتح عثمان): اللمع في العربية، تحقيق الدكتور سميح أبو مغلي، دار مجدلاوي للنشر عمان، ١٩٨٨م.
- ابن خلدون (ولي الدين عبد الرحمن بن محمد): مقدمة ابن خلدون، تحقيق عبد الله محمد الدرويش، دار البلخي - دمشق، ط١، الجزء الثاني، ٢٠٠٤م.
- الأنصاري (ابن هشام): مغني اللبيب عن كتب الأعاريب، تحقيق محمد محيي الدين عبد الحميد، المكتبة العصرية - بيروت، ١٩٩١م.
- أنيس (إبراهيم): من أسرار اللغة، مكتبة الأنجلو المصرية، ط٦، ١٩٧٨م.
- برينكر (كلاوس): التحليل اللغوي للنص، مدخل إلى المفاهيم الأساسية والمناهج، ترجمة: أ.د سعيد حسن البحيري، مؤسسة المختار للنشر والتوزيع، ٢٠١٠م.
- بحيري(سعيد حسن): نظرية التبعية في التحليل النحوي،مكتبة الأنجلو المصرية، ط١، ١٩٨٨م.

• البهنساوي (حسام): القواعد التحويلية في ديوان حاتم الطائي، مكتبة الثقافة الدينية، القاهرة ، ١٩٩٢م.
• البنهساوي (حسام): نظرية النحو الكلي والتراكيب اللغوية العربية، مكتبة الثقافة الدينية، ٢٠٠٤م.
• بونيه (الان): الذكاء الاصطناعي واقعه ومستقبله، ترجمة د.علي صبري فرغلي، عالم المعرفة، ١٩٩٣م.
• بوول (جيفري): النظرية النحوية، ترجمة د.مرتضى باقر، المنظمة العربية للترجمة، ط١، ٢٠٠٩م.
• الجرجاني (عبد القاهر): دلائل الإعجاز، تحقيق أبي فهر محمود محمد شاكر، مكتبة الخانجي - القاهرة، ط١، ١٩٩١م.
• الجرجاني (عبد القاهر): كتاب المقتصد في شرح الإيضاح، تحقيق د. كاظم بحر المرجان، دار الرشيد للنشر - الجمهورية العراقية ١٩٨٢م.
• حسان (تمام): مناهج البحث في اللغة، مكتبة الأنجلو القاهرة، ط١، ١٩٥٥م.
• حسان (تمام): مناهج البحث اللغوي، مكتبة الأنجلو المصرية، ١٩٩٠م.
• حسان (تمام): اللغة العربية معناها ومبناها، دار الثقافة، ١٩٩٤م.
• حميدة (مصطفى): نظام الارتباط والربط في تركيب الجملة العربية، مكتبة لبنان- ناشرون، ١٩٩٧م.
• خضر (محمد ذكي): الحروف العربية والحاسوب، مجلة مجمع اللغة الأردني-عمان، ١٩٩٦م.
• خليل (شرف الدين): الإحصاء الوصفي، شبكة الأبحاث والدراسات الاقتصادية، د.ت.
• خليل (حلمي): الكلمة دراسة لغوية معجمية، دار المعرفة الجامعية، الأسكندرية، ١٩٩٨م.
• الدحداح (أنطوان): معجم لغة النحو العربي، مكتبة لبنان ناشرون، ٢٠٠٠م.
• دعبول (موفق)، كيوان (فايز)، البواب (مروان)، وآخرون: معجم مصطلحات المعلوماتية (عربي - إنجليزي)، الجمعية العلمية السورية المعلوماتية، دمشق، ط١، ٢٠٠١م.

• الراجحي (عبده): النحو العربي والدرس الحديث "بحث في المنهج"، دار النهضة العربية للطباعة والنشر، ١٩٧٩م.

• زفارت (جوست): البنيات التركيبية والبنيات الدلالية، علاقة الشكل بالمعنى. ترجمة د.عبد الواحد خيري، ط١ ، دار الحوار للنشر والتوزيع، سوريا، ٢٠٠٨م.

• زكريا، (ميشال): الألسنية التوليدية وقواعد اللغة العربية (النظرية الألسنية)، ط٢، المؤسسة الجامعية للدراسات والنشر والتوزيع، ١٩٨٦م.

• الساقي (مصطفى): أقسام الكلام العربي من حيث الشكل والوظيفة، مكتبة الخانجي بالقاهرة، ١٩٧٧م.

• السعيد (المعتز بالله): مدونة معجم تاريخي للغة العربية «معالجة لغوية حاسوبية» أطروحة دكتوراه ، دار العلوم جامعة القاهرة، ٢٠١٠م.

• السيوطي (الإمام جلال الدين): همع الهوامع في شرح جمع الجوامع، تحقيق د.عبد العال سالم مكرم، مؤسسة الرسالة، ١٩٩٢م.

• سيبويه (أبو بشر عمرو بن عثمان بن قنبر): الكتاب (كتاب سيبويه)، تحقيق عبد السلام محمد هارون، مكتبة الخانجي بالقاهرة، ١٩٨٨م.

• شريف (عمرو): ثم صار المخ عقلا، طبعة مكتبة الشروق الدولية، ط ٢، ٢٠١٣م.

• شمس الدين(جلال): الأنماط الشكلية لكلام العرب، نظرية وتطبيقا دراسة بنيوية، مؤسسة الثقافة الجامعية، الأسكندرية، ط١، ١٩٩٥م.

• عاشور(المنصف): بنية الجملة العربية بين التحليل والنظرية، منشورات كلية الآداب بمنوبة، ١٩٩١م.

• عبادة (محمد إبراهيم): الجملة العربية، مكوناتها-أنواعها-تحليلها، مكتبة الآداب القاهرة، ط ٤، ٢٠٠٧م.

• عبد العزيز (محمد حسن):لغة الصحافة المعاصرة، دار المعارف، د.ت.

• عفيفي (أحمد مصطفى):النظرية النحوية-المفاهيم والتحديات، وقائع مؤتمر"العربية وقرن من الدرس النحوي" دار العلوم-القاهرة" ٢٠٠٣م.

• علي، (نبيل)، حجازي، (نادية): الفجوة الرقمية "رؤية عربية لمجتمع المعرفة"، عالم المعرفة، ٢٠٠٥م.

• على (نبيل):اللغة العربية والحاسوب، تعريب، ١٩٨٨م.

• غاليم (محمد): هندسة التوازي النحوي وبنية الذهن المعرفية. كتاب آفاق اللسانيات (تكريما للأستاذ الدكتور نهاد الموسى)، مركز دراسات الوحدة العربية، بيروت، طـ ١، ٢٠١١م.

• الفهري (عبد القادر الفاسي): البناء الموازي، نظرية في بناء الكلمة وبناء الجملة، دار توبقال للناشر، المغرب، ط ١، ١٩٩٠م.

• ابن منظور (محمد بن مكرم بن علي أبو الفضل جمال الدين): لسان العرب، طبعة القاهرة، دار المعارف، ١٩٨١م.

• مزوني (دومينيك): قيود النحو الكلي: كيف يكتسب البشر القدرة على الكلام، ترجمة فرحات المليح. مجلة الحياة الثقافية. السنة ٢٧. العدد ١٣٩، ٢٠٠٢م.

• مصلوح (سعد): الأسلوب دراسة لغوية إحصائية، عالم الكتب، ط٣، ١٩٩٢م.

• الموسى (نهاد): العربية نحو توصيف جديد في ضوء اللسانيات الحاسوبية، المؤسسة العربية للدراسات والنشر، ط١، ٢٠٠٠م.

• نواف (خالد حازم)، محمد (إبراهيم محمد): الصحافة الإلكترونية، ماهيتها والمسؤولية التقصيرية الناشئة عن نشاطها، مجلة الشريعة والقانون، العدد السادس والأربعين، ٢٠١١م.

ثانيًا: المراجع الأجنبية:

• Abeillé, A. (2003). Treebanks: Building and Using Parsed Corpora. Springer Science & Business Media.

• Arabic Tokenizer. [Version 3.6.0] [Computer Software] Stanford NLP group: Stanford.

• Aliwy, A. (2013). Arabic Morphosyntactic Raw Text Part of Speech Tagging System. PhD dissertation Under the Supervision of Prof. dr hab. Jerzy Tyszkiewicz, Faculty of Mathematics, Informatics and Mechanics, University of Warsaw.

• Alqrainy, sh. (2008). A Morphological – Syntactical Analysis Approach For Arabic Textual Tagging. PhD dissertation, School of Computing Faculty of Computing Sciences and Engineering De Montfort University.

• Annotald Program. [Version 1.0] [Computer Software] Github: Pennsylvania.

• Anton, H. & Rorres, Ch. (2010). Elementary Linear Algebra. John Wiley & Sons, Inc. 10th edition.

• Arnfield, S. (1996). Word Class Driven Synthesis of Prosodic Annotations. In Proceedings., Fourth International Conference on. Vol. 3. IEEE, 1996.

• Attardi, G. & Dell'Orletta, F. (2008). Chunking and Dependency Parsing. In Proc of LREC Workshop on Partial Parsing: Between Chunking and Deep Parsing. Marrakech, Morocco.

• Attia, M. (2004).Theory and Implementation Of a Large-Scale Arabic Phonetic Transcriptor, and Applications. PhD dissertation Under the Supervision of Prof.Mohsen Rashwan Faculty of Engineering, Cairo University.

• Attia, M. (2007). Arabic Tokenization Systems. In proceeding of ACL.

• ATKINS, B.T.S. (1992). Tools For Computer-Aided Corpus Lexicography: The HECTOR Project. Acta Linguistica Hungarica (An International Journal of Longuistics) Volume 41, Numbers 1-4.

• Babby, L. (2009). The Syntax of Argument Structure. Cambridge University Press.

• Badr, I. & Zbib, R. & Glass, J. (2008). Segmentation for English-to- Arabic Statistical Machine Translation. In Proceedings of ACL.

• Brants, T.& Skut, W. (1996). Syntactic Annotation Of A German Newspaper Corpus. In Proceedings of the ATALA Treebank Workshop, Paris, France.

• Brants, T. (1997). The NeGra Export For Annotated Corpora. Projekt C3 NebenLaufige Grammatische Verabeitung.

• Birbeck, M. et al., (2001). Professional XML 2nd Edition. Published by Wrox Press Ltd, the United States.

• Buckwalter, T. (2004). Buckwalter Arabic Morphological Analyzer Version 1.0. Lingistic Data Consortium, University of Pennsylvania, 2004. LDC Catalog NO: LDC2004L02.

• Bulk Rename Utility. [Version 3.0.01] [Computer Software] TGRMN Software.

• Cambria, E & Poria, S & Bisio, F & Bajpai, R & Chaturvedi, I. (2015). The CLSA Model: A Novel Framework for Concept-Level Sentiment Analysis. In Proceedings CICLing 2015, Egypt. Springer International Publishing Switzerland 2015. A. Gelbukh (Ed.).

• Cmerjrek, M. & Curin, J. & Havelka, J. (2006). Prague Czech-English Dependency Treebank. In Preceeding ACL Anthology.

• CRF++ [Version 0.7] [Computer Software] taku-ku: chasen

• C2D-v0.7. [Version 0.7] [Computer Software] New York, NY: Morningside.

• Diab, M.T. (2007). Improved Arabic Base Phrase Chunking with a new enriched POS tag set. In Proceedings of the 5th Workshop on Important Unresolved Matters, Association for Computational Linguistics (ACL), Prague.

• Diab, M.T. (2007). Improved Arabic Base Phrase Chunking With a new enriched POS tag set. Proceedings of the 5th Workshop on Important Unresolved Matters, Prague, Czech Republic.

• Dinneen, F P. (1995). General Linguistics.Georgetown. University Press.

• Dirven, R. & Langacker, R. (1992). Grammar in Mind and Brain. Mouton de Gruyter, Berlin. New York.

• Dukes, K. & Buckwalter, T. (2010). A Dependency Treebank of the Quran using Traditional Arabic Grammar. 7th international conference on Informatics and Systems. Cairo, Egypt.

• Ethem, A. (2010). Introduction to Machine Learning.The MIT Press Cambridge, Massachusttes, London,England. Second Edition.

• evalC. [Version 1] [Computer Software] Sangati, F: Edinburgh.

• Fellbaum, C. (1998). WordNet: An Electronic Lexical Database. Cambridge, Mass: MIT Press.

• Garsid, R. & Leech, G. & McEnery, T. (2013). Corpus Annotation. Second Published by Routledge. New York USA.

• Gibbon, D. & Mertins, I. & Moore, R . (2000). Handbook of Multimodal and Spoken Dialogue Systems: Resources, Terminology and Product Evaluation. Springer Science & Business Media.

• George A. M. (1995). WORDNET: A lexical Database for English, the ACM.

• Ghahramani, Z. (2001). An Introduction to Hidden Markov Models and Bayesian Networks. International Journal of Pattern Recognition and Artificial Intelligence.

• Green, S & D.Manning, C. (2010). Better Arabic Parsing: Baselines, Evaluations, and Analysis. In Proceedings of the 23rd International Conference on Computational Linguistics (Coling). Beijing.

• Gurevych, I. & Biemann, C. & Zesch, T. (2013). Language Processing and Knowledge in the Web. 25th International Conference, GSCL 2013, Darmstadt, Germany, September 25-27, 2013. Proceedings.

• Gulbransen, D. (2002).Using XML.Second Edition.Que Publishing.

• Habash , N. (2010). Introduction to Arabic Natural Language Processing, A Publication in the Morgan & Claypool Publishers series.

• Habash, N. & Faraj, R. & Roth, R. (2009). Syntactic Annotation in the Columbia Arabic Treebank. In Proceeding of ELDA.

• Habash, N. & Roth, R. (2009). CATIB: The columbia Arabic Treebank.Proceedings of the ACL-IJCNLP 2009 Conference Short Papers.Suntec, Singapore.

• Haji ̌c, J.& Ciaramita, M. & Johansson, R. et al., (2009). The CoNLL-2009 Shared Task: Syntactic and Semantic Dependencies in Multiple Languages In Proceedings of the 13th CoNLL-2009, June 4-5 , Boulder, Colorado.

• Hale, J. & Callaway, F. & Feldman, E et al., (2014). Modeling Neural Correlates of Syntactic Structure Building. In AMLaP 2014-Poster, Scotland.

• Han, ch. & Eon-Suk Ko, Na-Rae. (2002).Penn Korean Treebank: Development and Evaluation. In Preceeding ACL Anthology.

• Holmboe, H. (2004).Nordisk Sporogteknologi, Nordic Language Technology. Museum Tusculanums forlag.

• INDURKHYA, N. & DAMERAU, F. (2010). Handbook of Natural Language Processing. Second Edition.CRC Press.Taylor & Francis Group. London.

• Jurafsky.D, Martin.J. (2007). Speech and Language Processing: An introduction to natural language processing. Second Edition.

• Kennedy, G. (1998). An Introduction to Corpus Linguistics. Longman.

• Khoja, Sh. (2001). APT: Arabic Part-of-speech Tagger. In Proceedings of the Student Workshop at the Second Meeting of the North American Chapter of the Association for Computational Linguistics (NAACL2001), Carnegie Mellon University, Pittsburgh, Pennsylvania.

• Khoja, Sh. (2001). An Arabic Tagset for the Morphosyntactic Tagging of Arabic. In Proceedings of Corpus Linguistics, Lancaster University, Lancaster, UK.

• Koopman, H. & Sportiche, D. & Stabler, E. (2003). An Introduction to Syntactic Analysis and Theory . UCLA.

• Kurohashi, S. (1996). Building Japanese Parsed Corpus. In Proceedings of the First International Conference on Language Resources and Evaluation.

• Kulick, S. (2010). Simultaneous Tokenization and Part-of-Speech Tagging for Arabic without a Morphological Analyzer. Proceedings of the ACL 2010 Conference Short Papers. Uppsala, Sweden.

• Kilgarriff, A and Rosenzweing, J. (2000). English SENSEVAL: Report and Results. In LREC.

• Lanfranchi, A. & O'Gorman, T. & Howard, A. & Gould, K. & Regan, M. (2002). Bracketing Biomedical Text: An Addendum to Penn Treebank II Guidelines. Institute of Cognitive Science, University of Colorado at Boulder.

• Levy, R & Andrew, G. (2006). Tregex and Tsurgeon: tools for querying and manipulating tree data structures. In preceding of 5th International Conference on Language Resources and Evaluation (LREC 2006).

• Lüdeling, A. & Kytö, A. (2008). Corpus linguistics An International Volum 1. walter de Gruyter GmbH& Co. KG. Berlin, Germany.

• Maamouri, M. & Bies, A. et al., (2011). Penn Arabic Treebank Guidelines. Linguistic Data Consortium, University of Pennsylvania.

• Maamouri, M. & Bies, A & Bukwalter, T & Mekki, W. (2004). The Penn Arabic Treebank: Building a Large-Scale Annotated Arabic Corpus. In Proceedings of NEMLAR Conference on Arabic Language Resources and Tools.

• Marneffe, de & Dozat, T. & Sliverira, N. et al., (2014). Universal Stanford Dependencies: A cross-linguistic Typology. In Proceedings of 9th International Conference on Language Resources and Evaluation (LREC 2014).

• MADA+TOKAN. [Version 1] [Computer Software] New York, NY: Morningside.

• McDonald, R. & Niver, J. et al., (2013). Universal Dependency Annotation For Multilingual Parsing. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics. Sofia, Bulgaria.

• McEnery, T. & HARDIE, A. (2012). Corpus Linguistics "Method, Theory and Practice". First Edition. Edinburgh U.

- McEnery, T. & Wilson, A. (2001). Corpus Linguistics "An Introduction". Second Edition. Edinburgh University press.
- Mitkov, R. (2005). The Oxford Handbook of Computational Linguistics (Oxford Handbooks). Oxford University Press.
- Moreno, A.& Lopez, S. et al., (1997). Developing A Syntactic Annotation Scheme And Tools For A Spanish Treebank. In Proceedings of Treebanks workshop - Journées ATALA sur les corpus annotés pour la syntaxe.
- Murphy, Kevin P. (2012). Machine Learning A Probabilistic Perspective. The MIT Press.London, England.
- Notepad++.[Version 9.6] [Computer Software] ecreate: Don Ho.
- Nugues, P.M. (2014). Language Processing with perl and prolog. second Edition.springer.
- Olive, J. & Christianson, C & McCary, J. (2011). Handbook of Natural Language Processing and Machine Translation. Springer.
- Otero, C. (1994). Noam Chomsky: Critical Assessments.TJ Press (Padstow) Ltd, Padstow, Cornwall, London.
- Peng, B & Li, J & Chen, J & Han, X & Xu, R & Wong, K. (2015). Trending Sentiment-Topic Detection on Twitter. In Proceedings CICLing 2015, Egypt. Springer International Publishing Switzerland 2015. A. Gelbukh (Ed.).
- Ponomareva N, Rosso P, Pla F, Molina A. (2007). Conditional random fields vs. hidden markov models in a biomedical named entity recognition task. In Proc. of Int. Conf. Recent Advances in Natural Language Processing, RANLP.
- Pustejovsky, J. & Stubbs, A. (2012). Natural Language Annotation For Machine Learning. Frist Edition.O'Reilly Media.

- Radford, A. (2009). An Introduction to English Sentence Structure, University of Essex.
- Rambow, O. (2010). The Simple Truth about Dependency and Phrase Structure Representations .Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the ACL, Los Angeles, California.
- Sakaki, Sh & Miura, Y & Ma, X & Hattori, K & Ohkuma, T. (2014). Twitter User Gender Inference Using Combined Analysis of Text and Image Processing. In Proceedings of the 25th International Conference on Computational Linguistics. Dublin, Ireland.
- Sampson, G. & McCarthy, D. (2005). Corpus Linguistics: Readings in a Widening Discipline. A&C Black.
- Stanford-parser. [Version 3.6.0] [Computer Software] Stanford NLP group: Stanford.
- Sawalha, M. (2011).Open-source Resources and Standards for Arabic Word Structure Analysis: Fine Grained Morphological Analysis of Arabic Text Corpora. Submitted in accordance with the requirements for the degree of Doctor of Philosophy From The University of Leeds School of Computing.
- Sharaf, A. & Atwell, ES. & Dukes, K. & Sawalha, M. & Al-Saif, A. & Sharoff, S. & Markert, K. & Al-Sulaiti, L. & Abu Shawar, B. & Abbas, N. & Roberts, A. (2010). Arabic and Quranic computational linguistics projects at the University of Leeds. In Proceedings of the workshop of Increasing Arabic Contents on the Web, organized by Arab League Educational, Cultural and Scientific Organization (ALECSO).

- Smrž, O. & Bielický, V. et al., (2008). Prague Arabic dependency treebank: A word on the million words. In Proceedings of the Workshop on Arabic and Local Languages (LREC) 2008.Marrakech, Morocco. European Language 2008.Marrakech, Morocco. European Language Resources Association.
- Sutton, Ch. & McCallum, A. (2010). An Introduction to Conditional Random Fields.Cornell University Library.
- TIGERRegistry. [Version 1] [Computer Software] Letzte Änderung: Universität Stuttgart.
- Treebank_converter. [Version 1] [Computer Software] Lund University: Lund.
- Tregex. [Version 3.6.0] [Computer Software] Stanford NLP group: Stanford.
- Tou Ng, H. & Beng Lee, H. (1996). Integrating Multiple Knowledge Sources to Disambiguate Word Sense: An Exemplar-Based Approach. In Proceedings of the 34th annual meeting on Association for Computational Linguistics. Association for Computational Linguistics.
- Van Halteren, H. (1999). Syntactic Wordclass Tagging. Springer-Science+Business Media.B.V. University of Nijmegen.
- Wissler , L. & Almashraee, M. & Monett, D. & Paschke, A. (2014). The Gold Standard in Corpus Annotation. In proceedings of the In Proceedings of the 5th IEEE Germany Students Conference 2014. Passau, Germany.
- Wright, David R. (2005). “Finite State Machines”. CSC215 Class Notes. Prof. David R. Wright website, N. Carolina State Univ.
- xue, N. & Xia, F. et al., (2003). The Penn Chinese Treebank: Phrase Structure Annotation of a Large Corpus. Journal Natural Language Engineering.

فهرس الموضوعات

فهرس الأشكال

فهرس الجداول

الاختصارات المستخدمة

المصطلح بالعربية	المصطلح بالإنجليزية	الاختصار
أقسام الكلام	Parts Of Speech	POS
أدوات تحليل المدونة النحوية	Tools for Syntactic Corpus Analysis	TOSCA
استرجاع المعلومات	Information Retrieval	IR
مشجَّرات اتخاذ القرار	Decision Trees	DT
الاعتماديات العالمية	Universal Dependencies	UD
آلات الحالة المحدودة	Finite-State Machines	FSM
البنك الاعتمادي للنص القرآني	Quranic Arabic Dependency Treebank	QADT
بنك براغ الاعتمادي	Prague dependency Treebank	PDTB
بنك بنسلفانيا الشجري	Penn Treebank	PTB

المصطلح بالعربية	المصطلح بالإنجليزية	الاختصار
البنك الشجري	Treebank	TB
البنك الشجري النحوي-الدلالي للغة الإيطالية	Italian Syntactic-Semantic Treebank	ISST
بنك كولومبيا الشجري للغة العربية	The Columbia Arabic Treebank	CATiB
البنية الاعتمادية	Dependency Structure	DS
تحويل النص المكتوب إلى منطوق	Text to speech	TTS
الترجمة الآلية	Machine Translation	MT
الترجمة الآلية الإحصائية	Statistical Machine Translation	SMT
التعبيرات النمطية	Regular expression	Regex
التعرف على كينونة الاسم	Named Entity Recognation	NER
تعريفات نوع الوثيقة	Document Type Definitions	DTD
التعلم الآلي	Machine Learning	ML
تغطية للاستعمالات الإنجليزية	The Survey Of English Usage	SUE
جمعية اللسانيات الحاسوبية	The Association for Computational Linguistics	ACL
جواب السؤال	Question Answering	QA
الحقول العشوائية المشروطة	Conditional Random Fields	CRF
خوارزمية التحليل من أسفل لأعلى	Cocke–Younger–Kasami algorithm	CYK

المصطلح بالعربية	المصطلح بالإنجليزية	الاختصار
النحو الاحتمالي المتحرر من السياق	Probabilistic Context-Free Grammars	PCFG
الذكاء الاصطناعي	Artificial Intelligence	AI
الربط الخطابي	discourse connective	DC
رخصة جنو العمومية	GNU General Public License	GPL
شبكة البيانات اللغوية	Linguistic Data Consortium	LDC
شبكة كلمات برنستون	WordNet Princeton	PWN
صيغة تحويل نظام الحروف الدولي الموحّد بقوة ٨ بت	Unicode Transformation Format bit-8	UTF-8
صيغة تشومسكي العادية	Chomsky normal form	CNF
صيغة صفحة التشفير ١٢٥٦	Code page ١٢٥٦	CP-١٢٥٦
علم اللغة الحاسوبي	Computational Linguistics	CL
الفئة الكلامية المقلصة	Reduced Tagset	RTS
فك اللبس الدلالي للكلمات	Word Sense Disambiguation	WSD
فهم اللغة الطبيعية	Natural Language Understanding	NLU
لغة الترميز القابلة للامتداد	Extensible Markup Language	XML
المدونة الإنجليزية المنطوقة	Spoken English Corpus	SEC
المدونة التاريخية الموصفة للغة الآيسلندية	Icelandic Parsed Historical Corpus	IcePaHC
مدونة لند– لندن	London-Lund Corpus	LLC

المصطلح بالعربية	المصطلح بالإنجليزية	الاختصار
المدونة المعنونة نحويا	Syntactically Annotated Corpus	SAC
المدونة الوطنية البريطانية	The British National Corpus	BNC
مركز أنظمة التَّعلُّم الحاسوبيِّ	Center for Computational Learning Systems	CCLS
محلل باكولتر الصرفي للغة العربية	Buckwalter Arabic Morphological Analyzer	BAMA
معالجة اللغة الطبيعية	Natural Language Processing	NLP
معنون السمات الصرف-نحوية للعربية	Arabic Morphosyntactic Tagger	AMT
معهد ماساتشوستس للتقنية	Massachusetts Institute of Technology	MIT
المنظمة الدولية للمعايير القياسية	International Organization for .Standardization	ISO
منظمة العلوم والدفاع	Defence Science Organisation	DSO
مؤتمر تعليم اللغة الطبيعية	The Conference on Natural Language Learning	CoNLL
النحو الكلي	Universal Grammar	UG
نماذج ماركوف الخفية	Hidden Markov Models	HMM
وحدة أبحاث حوسبة اللغة الإنجليزية	Unit for Computer Research on the English Language	UCREL
وسيط	Argument	Arg

الرسائل الجامعية ٤

البنك الشجري النحوي

يعمل مركز الملك عبدالله بن عبدالعزيز الدولي لخدمة اللغة العربية على تعزيز خدماته في المجالات المتنوعة لخدمة اللغة العربية وعلومها، إذ ينطلق من رؤية موحّدة في أعماله عامة – ومنها برنامج النشر – وذلك بأن يطلق برامجه ودراساته في المجالات التي تفتقر إلى جهود نوعية، أو التي تحتاج إلى تكثيف العمل فيها.

ويجتهد المركز في انتقاء الكتب التي تصدر ضمن هذه السلسلة، بأن تكون مضيفة إلى حقلها المعرفي، ومفتاحا للمشروعات العلمية والعملية، ومحققة لتراكم معرفيٍّ مثرٍ.

وإذ تشيد الأمانة العامة في المركز بجهدٍ مؤلف الكتاب، تأليفاً، وتصحيحاً لمسوداته، ومراجعةً للطباعة، فإنها تدعو الباحثين كافة من أنحاء العالم إلى المساهمة في هذه السلسة، لتتكامل مع سلاسل المركز العلمية الأخرى.

ويسعد المركز بالعمل مع المؤسسات والأفراد المختصين والمهتمين في خدمة لغتنا العربية، وتكثيف الجهود والتكامل نحو تمكين لغتنا، وتحقيق وجودها السامي في مجالات الحياة.





ص.ب ١٢٥٠٠ الرياض ١١٤٧٣
هاتف:٠٠٩٦٦١١٢٥٨٧٢٦٨ – ٠٠٩٦٦١١٢٥٨١٠٨٢
البريد الإليكتروني: nashr@kaica.org.sa




9 786039 089551